N° 81-595-MIF au catalogue — N° 008
ISSN: 1704-8893
ISBN: 0-662-89631-9

Document de recherche

Éducation, compétences et apprentissage
Documents de recherche

L'incidence de l'éducation et de la formation des adultes sur la situation sur le marché du travail au Canada

par Shek-wai Hui et Jeffrey Smith

Division de la Culture, tourisme et centre de la statistique de l'éducation
2001 Immeuble principal, Ottawa, K1A 0T6
Téléphone : 1 800 307-3382 Télécopieur : 1 613 951-9040

Toutes les opinions émises par les auteurs de ce document ne reflètent pas nécessairement celles de Statistique Canada

Statistique Canada
Développement des ressources humaines Canada

Statistics Canada
Human Resources Development Canada

Canada

Éducation, compétences et apprentissage
Documents de recherche

L'incidence de l'éducation et de la formation des adultes sur la situation sur le marché du travail au Canada

Shek-wai Hui
Department of Economics, University of Winnipeg

Jeffrey Smith
Department of Economics, University of Maryland

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2003

Tous droits réservés. Il est interdit de reproduire ou de transmettre le contenu de la présente publication, sous quelque forme ou par quelque moyen que ce soit, enregistrement sur support magnétique, reproduction électronique, mécanique, photographique, ou autre, ou de l'emmagasiner dans un système de recouvrement, sans l'autorisation écrite préalable des Services de concession des droits de licence, Division du marketing, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Octobre 2003

N° 81-595-MIF2003008 au catalogue

Périodicité : Irrégulier

ISSN 1704-8893
ISBN 0-662-89631-9

Ottawa

This publication is available in English upon request (Catalogue no. 81-595-MIE2003008).

Statistique Canada
Développement des ressources humaines Canada

Les opinions exprimées par les auteurs de ce document ne reflètent pas nécessairement celles de Statistique Canada ou de Développement des ressources humaines Canada.

105-03939/

Comment obtenir d'autres renseignements

Toute demande de renseignements au sujet du présent produit ou au sujet de statistiques ou de services connexes doit être adressée à : Services aux clients, Culture, tourisme et centre de la statistique de l'éducation, Statistique Canada, Ottawa, Ontario, K1A 0T6; téléphone : (613) 951-7608; sans frais : 1 800 307-3382; télécopieur : (613) 951-9040; ou courrier électronique : educationstats@statcan.ca.

Pour obtenir des renseignements sur l'ensemble des données de Statistique Canada qui sont disponibles, veuillez composer l'un des numéros sans frais suivants. Vous pouvez également communiquer avec nous par courriel ou visiter notre site Web.

Service national de renseignements	1 800 263-1136
Service national d'appareils de télécommunications pour les malentendants	1 800 363-7629
Renseignements par courriel	infostats@statcan.ca
Site Web	www.statcan.ca

Renseignements sur les commandes

On peut se procurer ce produit n° 81-595-MIF2003008 au catalogue sur internet gratuitement. Pour obtenir un numéro de ce produit, les utilisateurs sont priés de se rendre à http://www.statcan.ca/cgi-bin/downpub/studiesfree_f.cgi.

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois, et ce, dans la langue officielle de leur choix. À cet égard, notre organisme s'est doté de normes de service à la clientèle qui doivent être observées par les employés lorsqu'ils offrent des services à la clientèle. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1 800 263-1136.

Remerciements

Nous aimerions remercier Développement des ressources humaines Canada de son appui financier et Statistique Canada de son soutien en matière d'accès aux données. Nous voulons exprimer notre reconnaissance à Lucie Gilbert pour son assistance, ses encouragements et sa patience tout au long de la préparation du présent rapport. Nous tenons aussi à remercier de leurs commentaires judicieux les participants à l'atelier sur l'Enquête sur l'éducation et sur la formation des adultes, tenu en mai 2001 à Ottawa.

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

Table des matières

Résumé		5
1.	Introduction	6
2.	Données	10
	2.1 Données de l'EEFA	10
	2.2 Définition des mesures de la formation suivie	11
	2.3 Définition des variables de résultats	13
3.	Évaluation de l'incidence de l'éducation et de la formation sur la situation sur le marché du travail	15
	3.1 Modèle des résultats sur le marché du travail en regard de la participation à la formation	16
	3.2 Paramètres à l'étude	19
	3.3 Méthodes d'évaluation transversales axées sur la sélection en fonction de facteurs observables	20
	3.4 Méthodes d'évaluation transversales fondées sur la sélection en fonction de facteurs non observables	24
	3.5 Hypothèses et données	26
4.	Estimations de l'incidence de la formation sur la situation sur le marché du travail	28
	4.1 Incidence de la formation sur l'emploi	28
	4.2 Gains hebdomadaires habituels	31
	4.3 Vraisemblance des estimations de l'incidence	32
	4.4 Sommaire des estimations de l'incidence	33
5.	Problèmes relatifs à l'EEFA aux fins de l'évaluation de l'incidence	35
6.	Conclusions	36
Tableaux		38
Références		47
Notes		51
Tableaux d'annexe		52
Culture, tourisme et Centre de la statistique de l'éducation		58
Éducation, compétences et apprentissage – Documents de recherche		59

N° 81-595-MIF2003008 au catalogue

Résumé

Dans ce rapport, nous nous appuyons sur les données de l'Enquête sur l'éducation et sur la formation des adultes (EEFA) de 1998 pour estimer l'incidence de la participation à l'éducation et à la formation des adultes sur l'emploi et les gains des Canadiens. Nous avons utilisé des méthodes axées sur « la sélection en fonction des facteurs observables », y compris les méthodes classiques de régression et les méthodes d'appariement par score de propension. Nous avons également appliqué des méthodes fondées sur des instruments ou restrictions d'exclusion, notamment l'estimation standard à l'aide de variables instrumentales et l'estimateur de sélection normal bivarié bien connu de Heckman. Ces méthodes permettent de traiter « la sélection en fonction de facteurs non observables ». Aucune des méthodes examinées n'a produit d'estimations plausibles de l'incidence de l'éducation et de la formation des adultes, bien que les méthodes axées sur la sélection en fonction de facteurs observables aient donné lieu à des estimations plus raisonnables que celles axées sur des instruments ou restrictions d'exclusion. À la lumière de nos résultats d'analyse, nous proposons des améliorations à l'EEFA qui feraient de l'enquête un outil mieux adapté à l'estimation de l'incidence de l'éducation et de la formation des adultes sur la situation sur le marché du travail.

1. Introduction

L'évaluation de l'incidence de l'éducation et de la formation des adultes est fort utile et ce, pour diverses raisons. Pour les décideurs, l'information relative aux effets des programmes d'éducation et de formation financés par l'État sur la situation sur le marché du travail présente un intérêt indéniable, une fois considérée dans un cadre coûts-bénéfices cohérent. Dans le même ordre d'idées, les renseignements sur les effets de l'éducation et de la formation des adultes autofinancées ou financées par l'employeur sur la situation sur le marché du travail donnent des indications quant à la capacité des politiques actuelles de subvention ou de taxation d'une telle formation (ou de l'augmentation des gains qui en découle, le cas échéant) d'influencer les décisions, potentiellement en deçà du niveau optimal, du point de vue de la société, de participation à ces activités. Pour les chercheurs, l'information sur les retombées de l'éducation et de la formation des adultes permet de cerner les modes d'accumulation du capital humain par les personnes et les entreprises et met en lumière les enjeux de l'économie politique, les contraintes de crédit susceptibles d'entraver les investissements individuels et collectifs optimaux dans le capital humain, et les théories de prestation insuffisante de services de formation sur le marché du travail.

On établit dans la littérature une distinction entre la formation offerte ou financée par l'État, particulièrement à l'intention des chômeurs et des travailleurs qui réintègrent le marché du travail, et celle offerte par les entreprises à leurs employés. Plusieurs facteurs expliquent cette distinction, notamment le fait que les groupes d'adultes qui reçoivent chacun de ces types d'éducation et de formation présentent des caractéristiques bien différentes et le fait que le contenu et la durée de la formation tendent aussi à différer considérablement. Nous tenons compte de cette distinction dans nos travaux empiriques présentés plus loin et dans la brève recension des écrits qui suit.

L'incidence des programmes gouvernementaux d'emploi et de formation sur la situation sur le marché du travail fait l'objet de nombreux ouvrages. Le tableau 25 du chapitre de Heckman, LaLonde et Smith (1999) dénombre des douzaines d'études réalisées dans plusieurs pays partout dans le monde. Bon nombre d'études américaines sont fondées sur des méthodes d'assignation aléatoire visant à produire des estimations crédibles, de grande qualité, de l'incidence des programmes d'aide à la recherche d'emploi, de formation en classe et de subventions salariales. Le tableau 22 du chapitre de Heckman, LaLonde et Smith (1999) présente une liste partielle de ces études. Au Canada, le nombre d'évaluations publiées de tels programmes est nettement plus limité. Deux ouvrages méritent cependant d'être soulignés : Park et al. (1996) et le Projet d'autosuffisance, largement cité, dont on retrouve un résumé dans Michalopoulos et al. (2000). Riddell (1991), Smith et Sweetman (2001) ainsi que Warburton et Warburton (2002) analysent et critiquent l'évaluation des programmes publics d'éducation et de formation des adultes au Canada. Heureusement, les programmes et les populations desservies aux États-

N° 81-595-MIF2003008 au catalogue

Unis sont assez semblables aux nôtres, de sorte que les évaluations américaines offrent des points de repère utiles auxquels il nous est possible de comparer les résultats de nos travaux.

La littérature sur les retombées de la formation fournie par l'employeur est bien plus mince. Plusieurs raisons expliquent ce phénomène. Premièrement, les administrations publiques sont plus enclines, et pour cause, à affecter des sommes considérables à l'évaluation de leurs propres programmes qu'à ceux des entreprises privées. Deuxièmement, les données valables sur la participation à la formation financée par l'employeur sont rares. Même dans les enquêtes d'envergure comportant des questions sur la formation financée par l'employeur, les erreurs de mesure posent des problèmes de taille[1].Troisièmement, si une formation financée par les administrations publiques est généralement suivie une seule fois ou à quelques occasions dans le cas des chômeurs, la formation financée par l'employeur s'étale souvent sur toute la vie professionnelle des employés. Les épisodes de formation sont généralement de courte durée, quelques jours ou quelques semaines, et on en compte souvent plusieurs par année. Comme on peut le voir ci-dessous, ces tendances caractérisent la formation financée par l'employeur au Canada. Ces caractéristiques de la formation financée par l'employeur font ressortir la nécessité de données longitudinales plutôt que transversales et rendent difficile le codage de la participation — selon la fréquence, le nombre d'heures, le nombre d'épisodes, etc. En Grande-Bretagne, Arulampalam, Booth et Elias (1997) ainsi que de Blundell, Dearden et Meghir (1996) ont tenté d'utiliser des données recueillies au moyen de panels pour étudier les effets de ce type de formation sur la situation sur le marché du travail et ont examiné ces questions en détail.

Heckman, Lochner, Smith et Taber (1997) de même que Carniero, Heckman et Manoli (2002) ont résumé les résultats de travaux visant à évaluer l'incidence des activités d'éducation et de formation des adultes financés par l'employeur sur la situation sur le marché du travail. Un résultat commun consiste en des impacts estimés élevés, généralement attribuable à l'incapacité des données disponibles de neutraliser complètement l'effet de la sélection présumée des employés les plus capables et les plus motivés pour les activités de formation de l'entreprise. Aucune étude de cette nature fondée sur des données canadiennes n'a été recensée.

Dans le présent rapport, nous estimons l'impact de la participation aux activités d'éducation et de formation des adultes à l'aide des données tirées de l'Enquête sur l'éducation et sur la formation des adultes (EEFA) de 1998. L'EEFA est un supplément de l'Enquête sur la population active (EPA) du Canada et comprend, par conséquent, tous les renseignements sur l'activité sur le marché du travail et les caractéristiques démographiques recueillis dans le cadre de l'EPA. Notre analyse vise deux principaux objectifs. Le premier est de commencer à combler les lacunes dans les écrits relativement à l'incidence de l'éducation et de la formation des adultes sur la situation sur le marché du travail au Canada. Le second consiste à déterminer la valeur de l'EEFA en tant que source de données permettant de mener des évaluations. Il ne s'agit pas là d'un objectif négligeable dans la mesure où l'EEFA vise principalement à documenter la nature et l'intensité de la participation aux activités d'éducation et de formation des adultes, ainsi qu'à permettre l'étude des facteurs déterminants de la participation à de telles activités.

Les écrits récents sur les méthodes économétriques permettant d'évaluer l'incidence de phénomènes comme l'éducation et la formation des adultes reposent sur une diversité de stratégies d'estimation[2]. Le problème fondamental auquel s'attachent tous ces estimateurs réside dans l'absence de données sur les traitements assignés de façon aléatoire. Faute d'assignation aléatoire, on peut se rabattre sur les données d'observation qui présentent l'inconvénient suivant : la variation observée dans le traitement (soit la variation observée au chapitre de la participation aux activités d'éducation et de formation des adultes, dans le cas qui nous intéresse ici) découle des choix d'optimisation (présumés) des personnes. Ces personnes disposent de renseignements dont ne dispose pas l'analyste et ont des caractéristiques que l'analyste ne peut observer. Par conséquent, les comparaisons simples de la situation sur le marché du travail des participants et des non-participants combinent les effets de la participation et les écarts attribuables à une participation non aléatoire. Ces écarts donnent lieu à un biais de sélection.

La littérature propose deux grandes catégories d'estimateurs pour résoudre ce problème : les estimateurs fondés sur l'hypothèse que les données sont suffisamment étayées pour qu'on puisse corriger la majeure partie des écarts systématiques entre participants et non-participants, et les estimateurs présumant l'absence de données de cette nature mais supposant la présence d'une variable (instrument ou restriction d'exclusion) ayant une incidence sur la participation mais non sur les résultats en l'absence de participation. Nous utilisons deux stratégies d'évaluation associées à chacune de ces grandes catégories. Dans la première catégorie, nous employons des méthodes classiques de régression, soit l'estimateur de Barnow, Cain et Goldberger (1980), de même que de nouvelles méthodes d'appariement par score de propension. Dans la seconde catégorie, nous appliquons les méthodes de variables instrumentales de même que l'estimateur normal bivarié de sélection bien connu de Heckman (1979). Dans chacun des cas, nos spécifications prennent appui sur ce que nous avons appris suite à notre analyse des facteurs déterminants de la participation aux activités d'éducation et de formation des adultes.

Comme l'observent notamment Heckman, LaLonde et Smith (1999) de même que Smith (2000), chacun de ces estimateurs pose des hypothèses différentes quant aux processus à l'origine de la participation aux activités d'éducation et de formation des adultes ainsi que de l'emploi et des gains, soit les deux facettes de la situation sur le marché du travail qui font l'objet de la présente étude. Tout au plus, les hypothèses de l'un des estimateurs considérés concordent avec les données et le contexte institutionnel examinés ici. Nous avons étudié tous les estimateurs dans l'espoir que les données nous mettent sur la voie des stratégies d'estimation les plus plausibles dans le cadre de l'EEFA et que les résultats nous permettent de proposer des améliorations aux données de l'EEFA pour les fins de l'estimation des impacts.

Les estimations de l'incidence obtenues à partir des méthodes économétriques que nous avons employées se sont révélées décevantes. Celles se rapportant à la formation financée par l'employeur sont nettement trop élevées, tandis que les estimations relatives à la formation financée par les administrations publiques sont souvent négatives. Les arguments théoriques fondés sur les taux de rendement prévus, de même que les comparaisons axées sur d'autres estimations provenant d'écrits qui prennent appui sur de meilleures données (souvent des données expérimentales dans le cas de la formation financée par l'État) jettent des doutes sérieux sur les résultats obtenus ici. Tous les estimateurs examinés ont produit des résultats

N° 81-595-MIF2003008 au catalogue

médiocres, mais ce sont l'estimateur de variables instrumentales et l'estimateur normal bivarié de sélection de Heckman qui ont donné les résultats les moins plausibles. Les analyses de sensibilité indiquent que ces résultats médiocres sont robustes lorsqu'on applique de faibles variations à la spécification estimée, ce qui nous porte à croire que le problème fondamental tient aux données plutôt qu'aux méthodes. En termes simples, les données de l'EEFA ne comprennent pas les éléments fondamentaux nécessaires à la production d'estimations plausibles de l'incidence de l'éducation et de la formation des adultes sur la situation sur le marché du travail.

Le reste de l'étude est structuré comme suit. À la section 2, nous décrivons les données de l'EEFA de 1998 utilisées ainsi que les mesures de la participation à la formation et des résultats sur le marché du travail. La section 3 fait état des méthodes d'estimation non expérimentales permettant d'estimer les effets de l'éducation et de la formation des adultes sur les gains et l'emploi des participants. La section 4 présente nos estimations de l'incidence et explique les problèmes relevés à la lumière des ouvrages théoriques et empiriques existants. À la section 5, nous résumons brièvement nos propositions destinées à faire de l'EEFA un meilleur outil d'estimation des impacts; ces suggestions sont décrites de façon plus détaillée dans notre article connexe, Hui et Smith (2003). Enfin, la section 6 présente un résumé des résultats et quelques conclusions relatives à l'éducation et à la formation des adultes au Canada.

2. Données

2.1 Données de l'EEFA

Les données employées sont tirées du fichier principal de l'Enquête sur l'éducation et sur la formation des adultes (EEFA) de 1998. L'EEFA de 1998 fut la troisième d'une série d'enquêtes comparables conçues pour mesurer la participation aux activités d'éducation et de formation des adultes, c'est-à-dire l'éducation et la formation reçues après la fin des études scolaires. L'enquête vise à mesurer les taux de participation, à déterminer le rôle des employeurs au chapitre de la participation à l'éducation et à la formation des adultes et de la prestation de ces activités, ainsi qu'à cerner les obstacles à l'éducation et à la formation des adultes. Statistique Canada a recueilli les données de l'EEFA pour le compte de Développement des ressources humaines Canada.

L'EEFA est un supplément de l'Enquête sur la population active (EPA), une enquête fondée sur un plan d'échantillon par panels chevauchants. Chaque mois, un nouvel échantillon aléatoire de la population de l'EPA — la population civile âgée de 15 ans et plus — est prélevé. Chacun de ces échantillons est appelé « groupe de renouvellement ». La taille des groupes de renouvellement est à peu près la même. Chaque groupe fait partie de l'échantillon de l'EPA durant six mois consécutifs, après quoi il n'est plus suivi et est remplacé par un nouveau. L'EEFA a été menée auprès de cinq des six groupes de renouvellement dans le cadre des Enquêtes sur la population active de janvier 1998 et de mars 1998[3].

L'EEFA de 1998 (ci-après appelée simplement EEFA) comporte cinq modules, nommés modules A à E. Les questions du module A permettent de recueillir des renseignements de base sur les répondants. On demande également aux répondants s'ils ont participé à une activité de formation ou d'éducation au cours de la dernière année. Ceux qui répondent par la négative sautent les trois modules suivants, B, C et D, et passent directement au module E.

Les questions du module B se rapportent à la formation et à l'éducation menant à des attestations d'études officielles. Dans l'EEFA, cette catégorie d'éducation et de formation est désignée par le terme « programmes de formation ». Les questions du module C recueillent des renseignements sur l'éducation et la formation ne menant pas à une attestation d'études officielle, ce qui dans l'EEFA est désigné par le terme « cours de formation ». Le module D porte sur les cours suivis à des fins de loisirs, de divertissement et d'enrichissement personnel ainsi que sur les activités de formation non déclarées dans les modules B et C. Puisque notre étude s'attache principalement aux effets de la formation sur la situation sur le marché du travail, nous n'avons pas tenu compte des cours dont les répondants ont fait état au module D. Dans chacun des trois modules B, C et D, l'enquête recueille des données sur un maximum de cinq cours ou programmes. Les données recueillies couvrent le domaine

N° 81-595-MIE2003008 au catalogue

d'études, le lieu, le fournisseur, les moyens d'éducation et la durée; elles permettent également de déterminer si la formation a été suivie ou non jusqu'à la fin, qui en a assumé les coûts et la nature du soutien de l'employeur (le cas échéant). L'enquête recueille également des données sur les raisons qui poussent les répondants à suivre une formation, sur les attentes qu'ils entretiennent à l'égard de la formation et sur leurs opinions quant à l'utilité de la formation. Toutes les questions des modules B, C et D renvoient aux activités de formation entreprises en 1997.

Les questions du module E s'adressent à tous les répondants. Ce module recueille des données relatives aux résultats sur le marché du travail en 1997, données qui ne sont pas fournies par l'EPA. Il s'agit notamment de renseignements sur l'emploi principal en 1997 si celui-ci diffère de l'emploi occupé au moment de la réalisation de l'EPA en 1998. Le module E recueille diverses données démographiques, y compris les caractéristiques des parents des répondants de même que la situation des répondants en regard de l'immigration et des handicaps. Pour compléter les données recueillies par le module E, on rattache les données sur le marché du travail tirées de l'EPA à l'enregistrement de chacun des répondants à l'EEFA. Le module E comprend en outre une série de questions servant à déterminer les raisons pour lesquelles des répondants n'ont pu participer à la formation qu'ils voulaient ou avaient besoin de suivre au cours de l'année de référence.

Le taux de réponse des ménages pour l'EPA s'est élevé à 95,3 %[4], et 85,2 % des répondants de l'EPA ont pris part à l'EEFA de 1998, de sorte que l'EEFA affiche globalement un taux de réponse de 81,2 %. L'échantillon de l'EEFA de 1998 comptait 33 410 répondants. Pour restreindre notre analyse à ceux qui avaient terminé leurs études scolaires, nous avons limité l'échantillon aux personnes âgées de 25 à 64 ans qui n'étudiaient pas à temps plein au moment de la réalisation de l'EPA. Le tableau 1 montre comment ces restrictions ont donné lieu à des échantillons d'analyse composés de 10 748 hommes et de 12 418 femmes. Les échantillons employés dans certaines analyses sont en fait plus petits compte tenu de la non-réponse à certaines questions sur des covariables particulières.

2.2 Définition des mesures de la formation suivie

Comme nous l'avons indiqué dans l'introduction, les écrits sur l'éducation et la formation des adultes (ci-après appelées simplement « formation ») établissent, pour diverses raisons, une distinction entre la formation d'ordre public et privé. Au Canada, les populations qui suivent une formation privée se différencient nettement de celles qui suivent une formation publique. Ces deux types de formation tendent également à différer sur le plan du contenu et de la nature des fournisseurs. Notre étude tient donc compte de cette distinction entre formation publique et formation privée.

L'EEFA distingue la formation soutenue par l'employeur et celle qui ne l'est pas. Dans l'EEFA, on interprète de façon très large le soutien de l'employeur, qui couvre par exemple les congés non payés pour fins de formation. Nous sommes toutefois d'avis qu'une personne à qui l'employeur a accordé un congé non payé pour suivre un programme de formation financé par l'État doit être considérée comme suivant une formation publique et non une formation privée.

Nous avons donc adopté une définition qui repose (nécessairement) sur les données disponibles dans l'EEFA, mais qui tient compte de l'entité assumant les coûts de formation. Nous avons établi à cette fin trois catégories s'excluant mutuellement : la formation financée par l'employeur ou le syndicat, la formation autofinancée et la formation financée par l'État ou d'autres organismes. La première catégorie regroupe la formation financée, en partie ou en totalité, par un employeur ou un syndicat. Cette catégorie prévaut sur les autres dans la mesure où toute formation financée à la fois par l'employeur et le gouvernement ou par l'employeur et l'employé est uniquement classée dans cette catégorie. La deuxième catégorie, celle de la formation autofinancée, couvre la formation payée exclusivement par le répondant de même que la formation dispensée gratuitement. Cette catégorie peut aussi inclure certaines activités de formation pour lesquelles l'État subventionne (en totalité ou en partie) le fournisseur de services, selon que les répondants tiennent compte ou non de ces subventions lorsqu'ils répondent à l'enquête. Elle peut également regrouper des activités de formation ayant fait l'objet de subventions indirectes prenant la forme de crédits d'impôt, d'aide au transport, d'allocations pour la garde d'enfants ou d'exemptions des obligations de recherche d'emploi. La dernière catégorie est en fait une classe résiduelle couvrant la formation qui, selon les répondants, est exclusivement financée par l'État ou par d'autres organismes. La grande majorité des activités de formation de la troisième catégorie sont considérées par les répondants comme étant exclusivement financées par l'État.

Comme nous l'avons déjà mentionné, l'EEFA établit une distinction entre « les programmes » et « les cours » selon que la formation mène ou non, ou est conçue pour mener ou non, à une attestation officielle. La participation à des programmes et la participation à des cours ne s'excluent pas mutuellement dans les données de l'EEFA, bien que la grande majorité des répondants aient suivi seulement l'un ou l'autre de ces types de formation. De plus, ils présentent certaines indications que des différences existent au chapitre des facteurs déterminants de la participation à des programmes et de la participation à des cours de formation. Cependant, compte tenu de la taille relativement restreinte de nos échantillons et du fait que cette distinction entre programmes et cours s'appuie sur le jugement des répondants, nous avons décidé de combiner ces deux types de formation dans notre définition des mesures de traitement.

Bien que certains répondants à l'EEFA aient déclaré avoir suivi plusieurs épisodes de formation en 1997, qu'il s'agisse de programmes ou de cours, la grande majorité d'entre eux n'en ont suivi qu'un. Par conséquent, dans notre analyse, les variables du « traitement » sont constituées des variables indicatrices (dichotomiques) relatives à la participation à une formation (cours, programmes ou les deux) financée par l'employeur ou le syndicat, à une formation autofinancée (cours, programmes ou les deux) ou à une formation (cours, programmes ou les deux) financée par l'État ou par d'autres organismes. Le tableau 2 présente des statistiques descriptives des trois mesures du traitement de même que les répartitions sous-jacentes des trois sources de financement selon qu'il s'agit de programmes ou de cours de formation.

Le tableau 3 présente des statistiques descriptives sur la répartition des heures consacrées aux épisodes de formation et donne ainsi une idée de l'intensité de la formation selon la catégorie. Le bloc supérieur indique la durée moyenne ainsi que le premier, le deuxième (médiane) et le troisième quartiles de la répartition pour l'échantillon global combinant les programmes et les cours. Les données sont

N° 81-595-MIF2003008 au catalogue

présentées séparément pour les hommes et les femmes et, dans ces catégories, elles sont présentées globalement et séparément selon la source de financement de la formation. Le tableau 3 permet de dégager quatre principales observations. Premièrement, la formation financée par l'employeur est généralement de plus courte durée que celle financée par l'État ou que la formation autofinancée. Cette observation s'applique tant pour les hommes que les femmes et aussi bien aux cours qu'aux programmes. Deuxièmement, comme on pouvait s'y attendre compte tenu des définitions adoptées, les programmes de formation tendent à être nettement plus longs que les cours de formation, bien que les deux répartitions se caractérisent par des chevauchements non négligeables. Troisièmement, on relève une similarité remarquable entre la durée de la formation chez les hommes et les femmes. Quatrièmement, les données révèlent une hétérogénéité très marquée de l'intensité du traitement dont nous cherchons à mesurer l'incidence. Par exemple, la formation financée par l'État pour les hommes a une durée moyenne de 591,5 heures. Toutefois, la valeur du premier quartile s'établit à 90 heures, tandis que celle du troisième quartile est supérieure à 1000 heures.

Un certain nombre d'épisodes de formation se poursuivaient au moment des entrevues de l'EEFA en 1998. Les personnes en cours de formation au moment de l'entrevue, soit au moment où les résultats sur le marché du travail sont mesurés, sont prises en compte dans les statistiques descriptives mais non dans les analyses de l'incidence.

2.3 Définition des variables de résultats

Les études de l'incidence de l'éducation et de la formation s'attachent généralement aux effets sur l'emploi et les gains. Nous nous intéressons à l'emploi parce que les personnes employées subviennent habituellement à leurs besoins plutôt que de dépendre de l'assurance-emploi ou de l'assistance sociale. Par conséquent, l'intégration au marché du travail est un objectif pour bon nombre de programmes gouvernementaux de formation et d'assistance. Parallèlement, pour les fins des analyses coûts-bénéfices traditionnelles, les gains représentent une mesure plus évidente des résultats. De plus, les gains reflètent les différences au chapitre des heures de travail et des taux de rémunération entre les différents emplois. Toutes choses étant égales par ailleurs, les administrations publiques (et les personnes suivant une formation) préfèrent que la formation financée ou subventionnée par l'État se traduise par des emplois à temps plein bien rémunérés plutôt que par des emplois à temps partiel associés à des taux salariaux inférieurs.

À l'instar des méthodes relevées dans la littérature, nous avons défini deux mesures de résultats, l'une liée à l'emploi, et l'autre aux gains. La première mesure des résultats est la situation d'activité du répondant au moment de l'entrevue menée dans le cadre de l'EPA en 1998. La seconde correspond aux gains hebdomadaires habituellement perçus par le répondant pour son emploi principal au moment de l'entrevue de l'EPA en 1998. Le tableau 4 indique la valeur moyenne de ces variables (et, dans le cas des gains, l'écart-type) pour les échantillons complets des hommes et des femmes, ainsi que selon la participation à chaque type de formation (financée par l'employeur ou le syndicat, autofinancée, et financée par l'État ou d'autres organismes) ou la non-participation à une formation en 1997.

Comme en discutent en détail Hui et Smith (2003), ces variables de résultats présentent un inconvénient de taille : elles sont mesurées pas plus de 12 mois (ou de 15 mois dans le cas des entrevues de mars 1998) après la fin de la formation dont nous cherchons à estimer les effets. Dans certains cas, l'intervalle peut être d'un mois ou deux seulement, alors que dans d'autres, la formation n'est peut-être même pas terminée. Par conséquent, les données ne reflètent pas intégralement l'incidence de la formation sur les gains ou l'emploi, particulièrement s'il faut un peu de temps pour trouver un emploi au terme de la formation. Des données récentes sur le programme GAIN de la Californie dont font état Hotz, Imbens et Klerman (2000) révèlent que l'acquisition de capital humain financée par l'État peut produire des bénéfices qui ne se manifestent que quelques années après la fin de la formation.

N° 81-595-MIF 2003008 au catalogue

3. Évaluation de l'incidence de l'éducation et de la formation sur la situation sur le marché du travail

Dans cette section, nous présentons un modèle des résultats obtenus sur le marché du travail en regard de la participation à la formation. Nous décrivons ensuite les hypothèses posées dans différentes méthodes économétriques d'estimation des effets de la formation sur la situation sur le marché du travail.

Le choix des méthodes d'évaluation est limité par le fait que l'EEFA est essentiellement une enquête transversale. L'EEFA recueille des renseignements auprès de chaque répondant une seule fois. Bien que l'EEFA fournisse des données sur les gains annuels totaux pour l'année 1997, soit la période pendant laquelle est suivie la formation que l'on analyse, cette mesure des gains n'est pas comparable à la mesure des gains hebdomadaires tirée de l'Enquête sur la population active de 1998. Par ailleurs, dans la plupart des stratégies d'estimation longitudinales, nous devons disposer de données sur la situation antérieure à la formation dont nous estimons l'incidence, plutôt que sur la situation au moment de cette formation. Le manque de renseignements précis sur le moment de la formation en 1997 restreint encore davantage les possibilités d'appliquer des méthodes longitudinales.

Par conséquent, la nature des données nous force à limiter notre analyse à des méthodes d'évaluation transversales. Nous considérons deux paires de méthodes connexes. La première paire se fonde sur l'hypothèse selon laquelle les données couvrent tous les facteurs importants ayant un effet tant sur la situation sur le marché du travail que sur la participation à la formation. La littérature désigne cette hypothèse par le terme « sélection en fonction de facteurs observables ». Nous examinons à la fois des estimateurs de régression paramétrique et des estimateurs d'appariement semi-paramétrique qui prennent appui sur cette hypothèse. La seconde paire de méthodes permet une sélection en fonction de facteurs non observables. Les deux méthodes exigent que les données comportent un instrument (ou restriction d'exclusion). Un instrument est une variable qui a une incidence sur la participation mais non sur les résultats, autrement que par son effet sur la participation. Les exemples crédibles de telles variables sont rares dans le contexte qui nous intéresse. Nous examinons la performance de plusieurs instruments potentiels dans les analyses empiriques présentées à la section 4. Comme les méthodes ne constituent pas l'objet principal de notre étude, nous les traitons brièvement en nous limitant à leurs éléments fondamentaux. Des renseignements plus détaillés sur toutes les méthodes examinées se retrouvent dans Angrist et Krueger (1999) ainsi que dans Heckman, LaLonde et Smith (1999).

3.1 Modèle des résultats sur le marché du travail en regard de la participation à la formation

L'équation de salaires standard du modèle de capital humain (voir par exemple Becker, 1964, ou Mincer, 1974) est le fondement des modèles de résultats que nous utilisons pour estimer l'incidence de la formation sur les gains et l'emploi à l'aide des données de l'EEFA. En supposant une forme fonctionnelle linéaire, nous avons l'équation de résultats

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + ... + \beta_K X_{Kit} + \delta_{i1} T_{i1t} + ... + \delta_{iJ} T_{iJt} + \varepsilon_{it},$$

où Y_{it} représente le résultat pour la personne « i » à la période « t » (les gains ou l'emploi dans le cas qui nous intéresse), où $X_{kit}, k = 1,...,K$ représentent les facteurs tels que le nombre d'années de scolarité et l'expérience, et où $T_{ijt}, j = 1,..,J$ sont des indicateurs de participation à divers types de formation. Comme nous ne disposons que de données transversales, nous laissons tomber l'indice « t ». Par souci de simplicité, nous définissons $Y_{1i} = Y_1(X_i, T_i = 1, \varepsilon_i)$ comme étant le résultat observé dans le cas de la participation à une formation et $Y_{0i} = Y_0(X_i, T_i = 0, \varepsilon_i)$ comme étant le résultat observé en l'absence de formation.

Examinons maintenant l'équation de participation. Pour simplifier les choses, supposons pour l'instant qu'il n'existe qu'un seul type de formation (de sorte que le choix se pose en termes de participation ou de non-participation à la formation) et que la formation n'est offerte qu'en une seule fois. Posons que $Y_i^*(X_i, T_i = 1)$ représente la valeur actualisée prévue des gains associés à la formation. De même, posons que $Y_i^*(X_i, T_i = 0)$ représente la valeur actualisée prévue des gains associés à la non-participation à une formation. Posons maintenant que $C(W_i)$ représente la valeur actualisée prévue des coûts associés à la formation, où W_i, qui peut comprendre des éléments de X_i, représente des facteurs faisant varier le coût de la formation d'une personne à l'autre. Il peut s'agir de facteurs tels que l'âge, le capital humain accumulé, les caractéristiques familiales, la branche d'activité, la profession, l'ancienneté professionnelle, la taille de l'entreprise, la région, etc. En supposant une fonction linéaire, nous obtenons la fonction des coûts de formation suivante :

$$C_i = \gamma_0 + \gamma_1 W_{1i} + ... + \gamma_L W_{Li} + u_i,$$

où $W_{1i},...,W_{Li}$ représentent les éléments individuels de W_i.

Une personne neutre au risque suivra une formation si et seulement si

$$Y_i^*(X_i, T_i = 1) - C(W_i) > Y_i^*(X_i, T_i = 0).$$

Les bénéfices (ou les pertes) actualisés prévus de la formation sont donnés par l'équation

$$H^*(X_i, W_i) = Y^*(X_i, T_i = 1) - Y^*(X_i, T_i = 0) - C(W_i).$$

Nous n'observons pas $H^*(X_i, W_i)$ parce que nous n'observons pas les gains contre-factuels prévus en l'absence de formation pour les personnes qui suivent une formation ni les gains contre-factuels prévus associés à la participation à la formation pour les personnes qui ne suivent pas une formation. C'est la décision de suivre ou

N° 81-595-MIF2003008 au catalogue

non une formation que les données nous permettent d'observer effectivement. Nous pouvons exprimer cette décision sous la forme d'un modèle de choix binaire, où $T_i = 1$ représente les personnes qui suivent une formation et $T_i = 0$, les personnes ne suivant pas une formation :

$$T_i = \begin{cases} 1 \; si \; H^*(X_i, W_i) > 0; \\ 0 \; autrement \end{cases}$$

En supposant que $Y^*(X_i, T_i)$ est une fonction linéaire de X_i et T_i (comme ci-dessus), que $C(W_i)$ est une fonction linéaire de W_i et que les éléments non observables des deux ont une distribution normale centrée à zéro, nous obtenons un modèle probit de la participation sous forme réduite.

La généralisation aux personnes qui ne sont pas neutres au risque est simple. Il suffit de remplacer les flux actualisés de gains ci-dessus par les flux actualisés d'utilité. La généralisation à de multiples types de formation est tout aussi simple, tant que nous continuons de supposer que la formation est offerte une seule fois — hypothèse qui cadre avec nos données transversales (mais non pas, bien entendu, avec la réalité). Dans ce cas, on obtient de multiples flux potentiels de gains ou d'utilité, l'un d'entre eux étant associé à l'absence de formation et les autres, à chaque type de formation disponible. Chaque personne choisit le parcours associé à la valeur actualisée optimale.

Examinons maintenant certaines répercussions de ce modèle simple de participation et de résultats sur l'estimation de l'incidence entreprise à l'aide de ce modèle. Il s'agit là d'un modèle de choix rationnel relativement à la participation à la formation. Les personnes suivent une formation lorsqu'elles s'attendent *ex ante* à ce que les bénéfices surpassent les coûts. Cette caractéristique du modèle a plusieurs conséquences. Premièrement, elle suppose une forte croyance au préalable que la formation, et notamment la formation privée (la formation financée par l'État étant parfois suivie pour d'autres raisons), aura une incidence positive sur la situation sur le marché du travail. Des incidences estimées négatives soulèveront des doutes et peuvent être considérées en quelque sorte comme un test informel de spécification.

La deuxième conséquence du comportement rationnel dans le contexte de ce modèle se rapporte à la sélection des variables instrumentales. Les personnes qui décident de suivre ou non une formation soupèsent les coûts et les bénéfices d'une telle décision. De bons instruments sont les variables qui ont une incidence sur les coûts et les bénéfices de la participation à une formation sans influencer les résultats en l'absence de formation. Citons parmi les instruments possibles selon cette logique les variables spécifiquement liées aux coûts, par exemple la distance par rapport au centre local de formation, et les variables liées à la variation des effets de la formation. La troisième conséquence du comportement rationnel se rapporte aux incidences hétérogènes. Si certaines personnes tirent de la formation plus d'avantages que d'autres (si tant est qu'elles en tirent), nous devrions nous attendre à ce que nos observations démontrent, à supposer que ces personnes puissent prévoir ces avantages dans une certaine mesure, des avantages plus marqués chez les personnes qui suivent une formation que chez les autres. Ces effets hétérogènes ont des conséquences importantes pour les politiques qui visent à favoriser la participation à la formation. On doit en effet faire preuve de prudence lorsqu'on cherche à

généraliser l'incidence estimée de la formation à des populations qui ne suivent pas de formation. Ces dernières ont peut-être des raisons d'agir ainsi.

Examinons maintenant la question du biais de sélection dans le contexte du modèle simple. Certaines variables exercent une influence tant sur la participation à la formation que sur les résultats en l'absence de formation. Si nous omettons de tenir correctement compte de ces variables lorsque nous estimons l'incidence de la formation, nos obtiendrons des estimations biaisées puisque l'indicateur de la formation se substituera aux variables manquantes qui conditionnent à la fois la formation et les résultats. Deux exemples types de telles variables sont la capacité et la motivation. La capacité comme la motivation ont vraisemblablement un effet positif sur les gains ainsi que sur la participation à la formation, ce qui implique un biais positif dans l'incidence estimée de la formation si ces deux variables interviennent vraiment et que nous omettons d'en tenir compte.

Si nous observons les variables pertinentes qui exercent une influence à la fois sur la participation à la formation et sur les résultats en l'absence d'une formation à partir de nos données, nous procédons à ce que Heckman et Robb (1985) appellent une « sélection en fonction des facteurs observables ». Dans ce cas, l'utilisation appropriée de ces variables à l'aide des méthodes décrites à la section 3.3 permettra de résoudre le problème de sélection. Si nous n'observons pas les variables pertinentes, alors, pour ce qui est du modèle, ces facteurs non observés donnent lieu à une corrélation entre les termes d'erreur des équations des résultats et de la participation, de sorte que $corr(\varepsilon_i, u_i) \neq 0$. Dans ce cas, nous devons adopter des méthodes de « sélection en fonction de facteurs non observables », méthodes que nous examinerons à la section 3.4. Ces méthodes nécessitent généralement un instrument, ou restriction d'exclusion, c'est-à-dire, dans notre modèle, une variable qui « appartient » à W_i mais non à X_i.

Passons enfin à la relation entre ce modèle simple et statique et le processus dynamique sous-jacent de la participation à la formation au cours du cycle de vie. Comme l'indique Becker (1964), il vaut mieux suivre une formation jeune plutôt qu'âgé, puisque les jeunes ont plus de temps pour voir se concrétiser les avantages que leur procure cette formation sur le marché du travail. Le modèle statique peut tenir compte de cette dimension dynamique de la décision de suivre une formation si l'on inclut l'âge parmi les facteurs déterminants de la formation.

La question du cycle de vie soulève également celle de la participation répétée à des activités de formation. Nous observons empiriquement des personnes qui suivent de la formation financée par l'État (voir par exemple Trott et Baj, 1993, pour les États-Unis) et de la formation privée (voir par exemple Blundell, Dearden et Meghir, 1996 pour le Royaume-Uni) à plusieurs reprises. Supposons que ces épisodes multiples de formation ne sont pas indépendants, mais qu'ils affichent plutôt une corrélation positive, peut-être en raison de différences non observées au chapitre des préférences pour la formation. Dans ce cas, l'incidence de la formation actuelle que nous estimons peut également inclure les effets des épisodes antérieurs de formation que nous n'observons pas. Dans la mesure où la formation a les conséquences positives que la théorie laisse entrevoir, nos estimations de l'incidence seront biaisées à la hausse si nous interprétons ces conséquences strictement comme des effets de la formation observée dans le cadre de l'EEFA.

N° 81-595-MIF2003008 au catalogue

3.2 Paramètres à l'étude

Dans un contexte où les personnes font l'expérience d'effets hétérogènes suite à la formation, il importe de considérer avec rigueur quel est le paramètre pertinent aux fins de l'évaluation de l'incidence de la formation[5]. Par souci de simplicité, nous supposons encore une fois qu'il n'existe qu'un type de formation ayant une incidence $\delta_i \ (= Y_{1i} - Y_{0i})$ pour la personne « i ». L'extension à des types multiples de formation est simple. Nous examinons trois paramètres possibles et discutons brièvement des rapports entre ceux-ci.

L'effet moyen du traitement (EMT) est, simultanément, l'effet de la formation sur une personne choisie au hasard dans la population visée et l'effet moyen sur l'ensemble de la population visée. Il se définit comme suit :

$$EMT = E(\delta_i).$$

Ce paramètre est pertinent dans les cas où une population sera tenue (ou persuadée) de suivre une formation.

Le paramètre de l'effet du traitement le plus courant dans la littérature est ce qu'on appelle le paramètre du « traitement sur le traité » (TT). Ce paramètre mesure l'incidence moyenne de la formation sur les personnes ayant suivi une formation selon les données. Dans la notation que nous avons adoptée, il est exprimé ainsi :

$$TT = E(\delta_i \mid T_i = 1).$$

Ce paramètre est pertinent si nous voulons effectuer une analyse coûts-bénéfices de la formation actuellement suivie, que celle-ci soit financée par des sources privées ou publiques.

Le dernier type de paramètre pertinent se compose de divers effets de traitement mesurés à la marge. Si nous utilisons un instrument binaire, nous pouvons définir les effets moyens locaux du traitement (EMLT), comme l'ont fait Imbens et Angrist (1994). L'EMLT est l'effet moyen sur les personnes dont la situation de participation change lorsque la valeur de l'instrument se modifie. Il suppose une réponse monotone, de sorte que les personnes ne deviennent pas, par exemple, plus susceptibles de participer si elles s'éloignent d'un centre de formation. Chaque instrument suppose son propre EMLT, et les EMLT de deux instruments distincts peuvent varier considérablement selon les impacts perçus par les personnes que chacun des instruments incite à participer. Si l'instrument est une variable de politique, comme les frais de formation exigés, l'EMLT peut présenter un grand intérêt du point de vue des politiques. Si nous utilisons un instrument continu, nous pouvons cerner l'effet du traitement marginal (ETM) comme dans Heckman et Vytlacil (2001). L'effet de traitement marginal, tel qu'ils le définissent, est l'effet sur la personne qui se montre indifférente à la formation compte tenu de la valeur actuelle de l'instrument, c'est-à-dire que la personne marginale est celle pour qui $H^*(X_i, W_i) = 0$. Heckman et Vytlacil (2001) montrent que tous les autres paramètres courants des effets du traitement peuvent être exprimés comme des intégrales particulières de tels ETM.

Enfin, nous pouvons définir d'autres effets marginaux qui ne sont pas forcément liés aux instruments. Si un nouveau centre de formation se construit dans une ville défavorisée, nous pouvons en cerner l'incidence sur les personnes qui choisissent de suivre une formation en raison de la construction du nouveau centre alors qu'elles n'en suivaient pas avant parce qu'elles devaient pour cela se rendre à la ville voisine. Cet effet du traitement est désigné par le terme « effet moyen marginal du traitement » (EMMT). Il ne s'agit pas d'un EMLT parce que la variation (le nouveau centre de formation) n'est pas un instrument, étant donné la localisation du centre dans une ville défavorisée où les résultats en l'absence de formation sont présumément moins élevés. Cette discussion montre bien l'intérêt que présentent les EMMT du point de vue des politiques.

Dans le présent article, nous ne nous attachons qu'à l'incidence du traitement sur la personne traitée pour différents types de formation et ce, pour diverses raisons. Premièrement, bien qu'il s'agisse du paramètre le plus souvent analysé dans la littérature, les estimations visant la formation financée par des sources publiques et celles visant la formation financée par des sources privées restent assez controversées, surtout dans le dernier cas. Deuxièmement, aucun des instruments que nous examinons ne se rapporte à une variation des politiques, ce qui est généralement nécessaire pour que les EMLT soient pertinents. Enfin, les EMT présentent relativement peu d'intérêt dans un contexte où personne ne propose de rendre obligatoire la formation publique ou privée.

3.3 Méthodes d'évaluation transversales axées sur la sélection en fonction de facteurs observables

Dans cette section, nous examinons deux méthodes axées sur la « sélection en fonction de facteurs observables », c'est-à-dire deux méthodes qui supposent l'observation, à partir des données, de tous les facteurs importants qui ont une incidence à la fois sur la participation à la formation et sur les résultats en l'absence d'une formation.

La méthode la plus fréquente (et, comme c'est souvent le cas, la plus simple) d'évaluation de l'incidence de la formation s'appuie sur les méthodes classiques de régression. Pour simplifier les choses, prenons d'abord le cas d'un type de formation et d'un effet commun de la formation. La généralisation à de multiples types de formation est simple; nous analyserons l'élargissement aux effets hétérogènes de la formation plus loin.

Supposons maintenant que X_i inclut les variables types qui font partie du modèle du capital humain, comme la scolarité et l'expérience (ou la variable substitutive de l'âge). Mais supposons que d'autres facteurs, qui ne sont pas inclus dans le modèle type du capital humain mais que l'on retrouve dans les données, ont une incidence à la fois sur les résultats en l'absence de formation et sur la participation à la formation. L'emplacement géographique en est un exemple possible. Ces dernières variables représentent un sous-ensemble de W_i. Posons que Z_i représente l'union de ce sous-ensemble de W_i et de X_i. Compte tenu de ces hypothèses, nous obtenons

$$E(\varepsilon_i \mid X_i, T_i) \neq 0 ,$$

N° 81-595-MIE2003008 au catalogue

mais

$$E(\varepsilon_i \mid Z_i, T_i) = 0. \qquad (1)$$

Barnow, Cain et Goldberger (1980) (ci-après appelés « BCG ») ont les premiers établi ces conditions pour estimer l'incidence de la formation à l'aide de méthodes classiques de régression, mais fondées sur un riche ensemble de covariables, suffisamment riche pour que le terme d'erreur de l'équation des résultats soit en moyenne conditionnellement indépendant de la formation.

Comme l'observent Heckman et Robb (1985) de même que Heckman et Smith (1996), dans un contexte d'effets hétérogènes, le terme d'erreur dans l'équation des résultats inclut maintenant implicitement la composante de l'incidence propre à la personne pour la population qui suit une formation. Cela signifie que le terme d'erreur comprend l'écart entre les effets moyens du traitement sur la personne traitée et les effets individuels de la formation propres à la personne en ayant suivi, ainsi que la composante non observée des résultats en l'absence de formation. Dans le modèle de BCG, le seul changement majeur réside dans l'interprétation. Désormais, le coefficient associé à l'indicateur de formation estime uniquement l'effet moyen du traitement sur la personne traitée; dans un contexte d'effets communs, ce coefficient est également une estimation de l'effet moyen du traitement.

À l'instar de l'estimateur de régression classique, l'appariement suppose une sélection en fonction des facteurs observables. Cependant, plutôt que postuler une équation de résultats sous une forme fonctionnelle, l'appariement compare directement les résultats des participants et des non-participants affichant les mêmes valeurs (ou des valeurs similaires) pour les variables considérées comme ayant une incidence tant sur la participation à la formation que sur les résultats en l'absence de formation.

L'appariement présente deux avantages et un inconvénient par rapport à l'estimateur de régression que l'on vient de décrire. Le principal avantage est que l'appariement est semi-paramétrique. Il n'est pas nécessaire de poser une hypothèse de forme fonctionnelle pour l'équation des résultats pour utiliser l'estimateur. Dans l'analyse de régression classique, même lorsqu'on emploie les covariables appropriées, on peut obtenir des estimations biaisées si on utilise une forme fonctionnelle incorrecte — en omettant, par exemple, d'inclure les termes requis d'ordre supérieur ou d'interaction. Le second avantage de cette méthode est qu'elle permet d'axer l'appariement sur des variables qui sont corrélées avec le terme d'erreur dans l'équation des résultats. Cela est possible parce que l'appariement exige uniquement que la moyenne du terme d'erreur soit la même pour les participants et les non-participants à la formation affichant des valeurs données pour les variables de conditionnement, et non qu'elle soit nulle. Dans notre notation, l'appariement exige que

$$E(\varepsilon_i \mid Z_i, T_i = 1) = E(\varepsilon_i \mid Z_i, T_i = 0),$$

mais il n'est pas nécessaire, comme dans le cas de la régression, que les deux termes équivalent à zéro. C'est là une hypothèse plus faible que l'hypothèse (1) ci-dessus[6]. L'appariement présente toutefois le désavantage suivant : si la forme fonctionnelle linéaire, implicite dans l'analyse de régression, cadre en fait avec les données, le fait

de ne pas l'imposer aura pour effet de réduire l'efficacité des estimations. En d'autres termes, si l'équation des résultats est effectivement linéaire, l'imposition de la linéarité produira des écarts-types moins élevés pour les estimations de l'incidence.

L'hypothèse d'indépendance conditionnelle (HIC) qui justifie l'appariement est exprimée ainsi :

$$Y_{0i} \perp T_i \mid Z_i .$$

Cette hypothèse suppose la condition d'équilibre mentionnée au paragraphe précédent. Selon l'HIC, conditionnellement à Z_i, les variables exerçant une influence à la fois sur la participation à la formation et sur les résultats en l'absence de formation, la participation à la formation n'est pas liée aux résultats en l'absence de formation. Autrement dit, peu importe le processus de sélection menant à la formation, dans les groupes caractérisés par les mêmes valeurs pour toutes les variables de Z_i, la participation n'est pas liée à ce qui adviendrait si la personne ne suivait pas de formation. Par conséquent, globalement, les participants à la formation peuvent avoir des perspectives meilleures ou pires, sur le marché du travail, que les non-participants mais, conditionnellement à Z_i, les résultats sur le marché du travail prévus des deux groupes sont équivalents. Deux points techniques méritent d'être soulignés. Premièrement, il s'agit de la version de l'HIC qui justifie le recours à l'appariement pour estimer l'incidence moyenne du traitement sur la personne traitée. Une version plus robuste, qui s'applique également à Y_{1i}, est nécessaire pour estimer les effets moyens du traitement. Deuxièmement, il se peut que les variables de Z_i ne soient pas des facteurs que peut changer le participant en prévision de sa participation à une formation (ou le non-participant en prévision de sa non-participation à une formation). Des analyses plus poussées se retrouvent dans Lechner et Miquel (2002) pour ce dernier point et dans Heckman, LaLonde et Smith (1999) pour le premier point.

L'appariement direct axé sur Z_i pose le problème suivant : tout ensemble de covariables qui cadre avec l'HIC de manière plausible sera d'une dimension relativement élevée. Même si tous les éléments de Z_i sont discrets, le nombre de combinaisons distinctes augmente très rapidement, ce qui cause un problème de cellules vides — valeurs de Z_i pour lesquelles nous observons des participants mais pas des non-participants qui permettraient d'estimer le contre-factuel. La solution qui consisterait simplement à ne pas tenir compte des cellules de participants sans non-participants correspondants n'est pas très satisfaisante. Sont tout aussi décevants les divers algorithmes ponctuels de combinaison de cellules utilisés dans certaines évaluations du programme de la *Comprehensive Employment and Training Administration* (CETA) aux États-Unis. Ces évaluations sont recensées et citées dans Barnow (1987).

Ce problème de l'absence potentielle de non-participants permettant de fournir le contre-factuel estimé des participants qui affichent certaines valeurs pour les variables de conditionnement est appelé « problème de support ». Le support est un terme statistique désignant l'ensemble des valeurs pour lesquelles la fonction de densité n'est pas nulle, soit l'ensemble des valeurs d'une variable pour laquelle on observe une probabilité positive. Outre l'HIC, la deuxième principale hypothèse sous-jacente à l'appariement est la condition de support, donnée par

N° 81-595-MIF2003008 au catalogue

$\Pr(T_i = 1 \mid Z_i) < 1$ pour toutes les valeurs possibles de Z_i. Cette condition veut qu'à toutes les valeurs des variables de conditionnement correspondent des non-participants. Même si cette condition s'applique dans la population, il peut arriver qu'elle ne soit pas satisfaite dans des échantillons de taille finie. On doit recourir à une variante plus générale de la condition de support pour estimer les effets moyens du traitement[7].

Les méthodes fondées sur la régression, comme l'estimateur de BCG, « règlent » implicitement le problème de support grâce à l'hypothèse de la forme fonctionnelle linéaire. L'hypothèse de la forme fonctionnelle comble les lacunes créées par les données manquantes. Cela révèle un autre avantage de l'appariement : l'appariement met en évidence la condition de support et permet de déterminer clairement si les résultats obtenus sont issus des données ou si, au contraire, les contre-factuels dépendent largement de l'hypothèse de linéarité.

On relève dans la littérature une convergence vers une nouvelle solution au « fléau de la dimensionnalité » et au problème connexe de support. Rosenbaum et Rubin (1983) ont montré que s'il est possible d'apparier en fonction de Z_i, c'est-à-dire si Z_i cadre avec l'HIC, il est également possible d'apparier en fonction de $P(Z_i) = \Pr(T_i = 1 \mid Z_i)$ Cette valeur représente la probabilité de participation ou « score de propension ». Cette méthode permet de résoudre le problème parce $P(Z_i)$ est un scalaire — simplement un nombre réel entre zéro et un, plutôt qu'un vecteur.

La littérature présente diverses méthodes d'application de l'appariement par score de propension. Parmi ces méthodes, on retrouve l'appariement selon le plus proche voisin (avec et sans substitution), l'appariement par cellules, l'appariement par noyaux et l'appariement local linéaire. Ces méthodes sont toutes convergentes au sens où, à mesure que la taille de l'échantillon augmente arbitrairement, elles produisent toutes le même résultat parce que dans un échantillon arbitrairement grand, toutes les méthodes se fondent uniquement sur les comparaisons entre participants et non-participants affichant des valeurs équivalentes de $P(Z_i)$. Des analyses détaillées des diverses méthodes sont présentées dans Heckman, Ichimura et Todd (1997), Heckman, LaLonde et Smith (1999) ainsi que Smith et Todd (2003).

Dans le présent article, nous nous intéressons à la substance plutôt qu'aux méthodes, de sorte que nous limiterons notre propos à l'appariement selon le plus proche voisin avec substitution, en faisant varier toutefois le nombre de plus proches voisins. Considérons d'abord un seul plus proche voisin. L'appariement selon le plus proche voisin sans substitution consiste à parcourir les observations des personnes traitées (participants) une à une et à trouver pour chaque observation le non-participant ayant le score de propension estimé le plus proche (en valeur absolue). Ce non-participant représente le plus proche voisin apparié au participant et ne peut être apparié à aucun autre participant. L'appariement selon le plus proche voisin avec substitution fonctionne de la même manière sauf qu'un non-participant peut être apparié à plus d'un participant. L'appariement avec substitution réduit la distance moyenne (dans les scores de propension) entre chaque participant et le non-participant auquel il est apparié. Cette méthode devrait réduire le biais. En revanche, l'appariement multiple de certains non-participants augmentera la variance de l'estimation produite. Dehejia et Wahba (1999) illustrent clairement le problème de l'appariement sans substitution lorsque le nombre d'observations de non-participants ayant de fortes probabilités de formation est inférieur au nombre de participants

ayant de fortes probabilités de formation (comme c'est généralement le cas pour des raisons évidentes). Dans cette situation, on obtient un nombre élevé de mauvais appariements. Pour éviter ce problème, il convient d'adopter la méthode de l'appariement avec substitution. La formule de l'estimateur du plus proche (et unique) voisin est la suivante :formule de l'estimateur du plus proche (et unique) voisin est la suivante :

$$\frac{1}{n_{\{T_i=1\}}} \sum_{i=1}^{n_{\{T_i=1\}}} (Y_{i1} - \sum_{j=1}^{n_{\{T_i=0\}}} w_{ij} Y_{0j}),$$

où la somme sur « i » se rapporte aux participants, la somme sur « j » se rapporte aux non-participants, $n_{\{T_i=1\}}$ représente le nombre de participants, $n_{\{T_i=0\}}$ représente le nombre de non-participants, et où

$$w_{ij} = \begin{cases} 1 \text{ si } j = \arg\min_j \{|P(Z_i) - P(Z_j)|\}, \\ 0 \text{ autrement.} \end{cases}$$

La généralisation aux cas de plus proches voisins multiples, chacun doté d'un poids égal, est directe.

La variation du nombre de plus proches voisins dans les estimations nous permet d'exercer un arbitrage entre le biais et la variance de notre estimateur. Examinons le passage d'un seul plus proche voisin pour établir le contre-factuel de chaque observation à deux plus proches voisins. La distance moyenne (exprimée par les scores de propension) entre chaque participant et les non-participants utilisés pour estimer la moyenne contre-factuelle du participant augmente nécessairement. Parallèlement, le nombre d'observations utilisées pour construire le contre-factuel augmente, ce qui réduit la variance de l'estimateur. Le nombre optimal dépend de la densité des non-participants. Par exemple, si le nombre de non-participants n'est pas nettement supérieur à celui des participants, le recours à des voisins supplémentaires présente peu d'intérêt. Nous avons fait des essais avec un, deux et cinq plus proches voisins dans nos travaux empiriques.

La norme relevée dans les ouvrages économiques traitant de l'appariement consiste à employer des méthodes « bootstrap » pour estimer les écarts-types, pour des raisons dont font notamment état Heckman, Ichimura et Todd (1997, 1998). Pour gagner du temps et faciliter les calculs, nous présentons ici des estimations qui ne tiennent pas compte des composantes de la variance issues de l'estimation des scores de propension ou de l'appariement lui-même. Par conséquent, nos écarts-types estimés sont vraisemblablement sous-évalués.

3.4 Méthodes d'évaluation transversales fondées sur la sélection en fonction de facteurs non observables

Nous examinons ici deux méthodes (connexes) qui s'appuient sur la sélection en fonction de facteurs non observables. Dans ce modèle, nous croyons que nos données ne fournissent pas toutes les variables qui influent à la fois sur la participation et sur les résultats en l'absence d'une participation. Nous pensons toutefois disposer d'une

variable, un instrument ou restriction d'exclusion, qui a une incidence sur la participation mais non sur les résultats autrement que par ses effets sur la participation.

Le premier de ces estimateurs est l'estimateur normal bivarié de sélection de Heckman (1979). Cet estimateur est fondé sur l'hypothèse selon laquelle les termes d'erreur dans les équations de la participation et des résultats suivent une distribution normale conjointe et selon laquelle le biais de sélection résulte d'une corrélation non nulle entre les deux termes d'erreur. Lorsque la variable des résultats est binaire, ce modèle correspond à un modèle probit bivarié. Techniquement, le modèle de Heckman (1979) est identifié uniquement par l'hypothèse de distribution normale conjointe et aucune restriction d'exclusion n'est requise. Des travaux poussés axés sur des analyses essais-erreurs et des études de Monte Carlo montrent qu'en pratique une restriction d'exclusion est essentielle à la stabilité du modèle. Les ouvrages publiés révèlent également que la performance de l'estimateur normal bivarié dépend fondamentalement de la validité de l'hypothèse de distribution normale. Les résultats de simulation présentés dans Heckman, LaLonde et Smith (1999) démontrent que cette performance dépend aussi du recours à une restriction d'exclusion solide — c'est-à-dire que la variable incluse dans l'équation de la participation, mais non dans celle des résultats, a une incidence suffisamment importante sur la participation. Puhani (2000) résume une grande partie des écrits méthodologiques traitant de la performance de l'estimateur normal bivarié.

Le modèle normal bivarié est souvent appelé « modèle à deux étapes de Heckman » en raison de la procédure simple d'estimation à deux étapes décrite dans Heckman (1979). Cependant, l'estimation en deux étapes n'est pas efficace. Étant donné que bon nombre de progiciels courants (Stata, par exemple) comprennent maintenant des routines permettant d'estimer conjointement les équations de la participation et des résultats, il convient de procéder ainsi et d'abandonner le nom attribué à l'estimateur. Il existe des versions du modèle de Heckman pour les effets communs et pour les effets hétérogènes. Björklund et Moffitt (1987) ont mis au point cette dernière version. Pour simplifier les choses, nous interprétons ici nos résultats en fonction du modèle des effets communs.

L'estimateur de variables instrumentales (VI) s'appuie lui aussi sur la sélection en fonction de facteurs non observables. En termes simples, l'estimateur des VI s'applique par une procédure des moindres carrés à deux étapes. À la première étape, on effectue une régression de la variable endogène, soit la participation à la formation dans le cas qui nous intéresse, en fonction des variables exogènes, comprenant à la fois les X_i de l'équation des résultats et l'instrument Z_i. Les estimations produites à partir de cette régression servent ensuite à générer les valeurs prévues de T_i. Contrairement aux valeurs initiales de T_i, ces valeurs prévues ont, par construction, une corrélation nulle avec le terme d'erreur, puisque toute la variation au chapitre de la participation est attribuable à la variation dans les variables exogènes X_i et Z_i. L'équation des résultats est alors estimée au moyen des valeurs prévues T_i de plutôt qu'en fonction des valeurs observées.

Comme l'indiquent Heckman et Robb (1985), l'estimateur des VI nécessite des hypothèses différentes de celles de l'estimateur normal bivarié et qui sont généralement plus faibles. Si la difficulté de trouver des variables instrumentales valides a dominé les débats sur cet estimateur dans la littérature, les ouvrages méthodologiques récents s'attachent à deux autres préoccupations. La première est

la force de l'instrument. Comme le montrent les simulations présentées dans Heckman, LaLonde et Smith (1999), des instruments faibles — c'est-à-dire des instruments n'ayant qu'une faible relation avec la participation à la formation — donnent lieu à des estimations très instables. Bound, Jaeger et Baker (1995) de même que les travaux qui s'en inspirent indiquent également que les estimations des VI fondées sur des instruments faibles peuvent être biaisées. À la lumière de ces recherches, les travaux empiriques accordent désormais plus d'attention à l'importance de l'instrument pour prédire la variable endogène.

La seconde préoccupation se rapporte à l'interprétation des estimations des VI dans un contexte d'effets hétérogènes du traitement. Un instrument peut être valide s'il n'est pas corrélé avec le terme d'erreur ε_i de l'équation des résultats; mais il peut néanmoins être corrélé avec la composante de l'incidence propre à la personne. Dans cette situation, l'interprétation de l'estimation produite par les variables instrumentales pose problème. Suivant certaines hypothèses, l'estimation produite peut renvoyer aux EMLT plutôt qu'à l'incidence du traitement sur la personne traitée, laquelle constitue généralement le paramètre pertinent. Heckman (1997) analyse cette question en détail.

3.5 Hypothèses et données

Le tableau 5 résume les variables comprises dans notre ensemble de données. Dans quelle mesure ces variables satisfont-elles aux exigences des estimateurs décrits aux sections 3.3 et 3.4? Pouvons-nous véritablement employer les estimateurs de BCG, d'appariement, de VI et l'estimateur normal bivarié, compte tenu de nos données et de ce que nous apprennent les travaux sur les déterminants de la participation à la formation et sur les résultats en l'absence de formation? Telles sont les questions abordées dans cette section.

Pour faciliter la compréhension, le tableau 6 répartit les variables figurant au tableau 5 entre trois ensembles, A, B et C. L'ensemble A regroupe les variables qui, selon nous, influent à la fois sur les résultats et sur la participation et qui sont vraisemblablement exogènes (non corrélées avec le terme d'erreur) dans l'équation des résultats. Cet ensemble couvre les variables types comprises dans les fonctions des gains du modèle de capital humain. Dans de tels modèles, l'éducation cerne les différences observées entre les personnes, au chapitre du capital humain, qui sont attribuables aux différences dans le niveau de scolarité atteint, tandis que l'expérience cerne les effets du capital humain accumulé grâce à l'apprentissage par la pratique. En l'absence de données sur le nombre d'années d'expérience, comme dans le cas de l'EEFA, l'âge tient lieu de variable substitutive.

En pratique, d'autres variables exerçant une influence sur la situation sur le marché du travail entrent aussi dans les fonctions de gains du modèle de capital humain (et dans les équations de l'emploi). Parmi de telles variables figurent les variables dichotomiques de la région qui cernent les différences d'ordre géographique quant aux conditions sur le marché du travail (et les niveaux de prix dans le cas des gains), de même qu'un indicateur de la situation d'immigration permettant de cerner les écarts attribuables aux compétences linguistiques, à la qualité de l'éducation ou à la discrimination. Nous adoptons ici la pratique standard qui consiste à estimer l'incidence séparément pour les hommes et les femmes plutôt que d'inclure une

N° 81-595-MIE2003008 au Catalogue

variable indicatrice pour le sexe. Des différences bien connues au chapitre de l'offre de travail selon le cycle de vie dues aux grossesses et à d'autres facteurs, ainsi que le rejet statistique fréquent des modèles qui regroupent les hommes et les femmes, justifient l'adoption de cette pratique. Il n'est donc pas nécessaire d'inclure la variable du sexe dans l'ensemble A. Nous intégrons les variables de l'ensemble A aux équations des résultats de tous les modèles fondés sur la régression que nous estimons et à toutes les spécifications du score de propension pour les estimations d'appariement.

L'ensemble B regroupe les variables qui, selon nous, ont une incidence tant sur la participation à la formation que sur les résultats en l'absence de formation, mais qui nous semblent endogènes (c'est-à-dire corrélées avec le terme d'erreur) dans l'équation des résultats. Ces variables couvrent principalement les caractéristiques de l'emploi en 1997, y compris les catégories relatives à la taille de l'entreprise et les indicateurs d'affiliation syndicale, d'emplois de professionnel et de col bleu. Comme certaines personnes de l'échantillon ne travaillaient pas en 1997, nous utilisons des variables indicatrices pour les valeurs manquantes relatives aux caractéristiques de l'emploi de même que des indicateurs de toute forme d'emploi et d'emploi à temps plein. Nos préoccupations quant à l'endogénéité nous empêchent d'inclure ces variables dans les modèles axés sur la régression, puisqu'une telle inclusion aurait vraisemblablement pour effet d'introduire un biais dans les estimations produites. Toutefois, dans la mesure où nous ne croyons pas qu'on puisse délibérément modifier ces variables en prévision d'une éventuelle participation à la formation, nous ne voyons pas d'inconvénient à les inclure dans certaines de nos spécifications d'appariement par score de propension.

L'ensemble C regroupe les variables qui, selon nous, peuvent servir d'instrument pour la participation à la formation. Ces variables couvrent diverses mesures relatives au conjoint du répondant (le cas échéant) et à ses enfants. La variable la plus problématique ici est la présence ou non d'un conjoint, certains travaux (controversés) laissant croire à un avantage salarial dans le cas des hommes mariés. Si un tel avantage existe véritablement, la situation de personne mariée ne constitue pas un bon instrument. Waite et Gallagher (2001) résument les écrits traitant de cette question. En raison des doutes que soulève ce phénomène, nous avons répété notre analyse après avoir éliminé de l'ensemble C les variables relatives à l'état matrimonial, ce qui nous a donné des résultats similaires sur le plan qualitatif. Nous utilisons les variables de l'ensemble C comme instruments pour les estimations de nos VI et comme restrictions d'exclusion pour nos estimations normales bivariées.

4. Estimations de l'incidence de la formation sur la situation sur le marché du travail

Nous présentons dans cette section nos estimations de l'incidence de la formation sur la situation sur le marché du travail à la lumière des données de l'EEFA. Nos observations sont réparties en fonction des deux variables de résultats qui nous intéressent ici, soit l'emploi et les gains. Un bref résumé des estimations de l'incidence conclut la section.

Nous examinons plusieurs modèles différents parce que les données de l'EEFA disponibles ne nous permettent pas de déterminer clairement *a priori* dans quelle mesure les hypothèses de chacun des modèles concordent avec la participation de la population et les processus de résultats pour ce qui est de l'éducation et de la formation des adultes au Canada. Comme l'indique notamment Smith (2000), aucun estimateur d'évaluation économétrique ne peut s'imposer comme le bon estimateur dans tous les cas, peu importent les données disponibles et le contexte à l'étude, bien qu'on ne puisse blâmer ceux qui tirent une telle conclusion d'une grande partie des ouvrages publiés. C'est pourquoi nous considérons ici des estimations tirées de plusieurs estimateurs différents.

4.1 Incidence de la formation sur l'emploi

Le tableau 7 présente nos estimations de l'incidence de la formation, quelle qu'elle soit, sur l'emploi, définies à partir des données de l'EPA de 1998. Pour faciliter la présentation, les estimations des autres paramètres de chaque modèle figurent en annexe (tableaux A.1 et A.2), où l'on retrouve également les estimations des modèles des scores de propension, de l'équation de la participation dans le modèle probit bivarié et le modèle normal bivarié, ainsi que les estimations résultant de la première étape du modèle à VI.

Chaque ligne du tableau 7 correspond à un estimateur différent, les estimateurs apparaissant dans l'ordre dans lequel ils ont été analysés à la section 3. Il convient d'interpréter toutes les estimations comme des estimations de l'incidence du traitement sur la personne traitée, comme on l'explique à la section 3.2. À la première ligne, on retrouve les estimations tirées du modèle de probabilité linéaire (MPL) inspiré de l'estimateur de BCG et produites à partir des variables de l'ensemble A. La deuxième ligne présente le même modèle, mais estimé à l'aide d'un probit. L'estimation de l'incidence correspond aux effets marginaux évalués en fonction de la moyenne des covariables, de sorte qu'elle est comparable à l'estimation produite à partir du MPL et des autres modèles.

Les six lignes suivantes donnent diverses estimations d'appariement par score de propension. Trois paires d'estimation d'appariement sont présentées. Chaque

N° 81-595-MIF2003008 au catalogue

paire s'appuie sur le même nombre de plus proches voisins mais fait varier l'ensemble de variables de conditionnement. Nous présentons des estimations fondées sur un, deux et cinq plus proches voisins. Pour chacune des paires, nous fournissons des estimations à partir des scores axés uniquement sur les variables de l'ensemble A et des scores axés sur les variables des ensembles A et B. Les comparaisons dans chacune des paires indiquent les effets éventuels du conditionnement sur un ensemble plus vaste de covariables — covariables qui ne pouvaient être intégrées aux modèles fondés sur la régression en raison de leur corrélation avec le terme d'erreur de l'équation des résultats.

Les deux dernières lignes présentent des estimations à partir de modèles qui tentent de tenir compte d'une sélection en fonction de facteurs non observables, la capacité ou la motivation par exemple. L'avant-dernière ligne du tableau présente les estimations des moindres carrés à deux étapes que nous avons obtenues en utilisant les variables de l'ensemble C comme instruments pour la formation dans une régression de l'emploi en fonction des variables de l'ensemble A plus l'indicateur de formation. Les écarts-types estimés de ces estimations tient compte de la variation supplémentaire attribuable à l'estimation de la première étape; l'ensemble complet des estimations de première étape (y compris les statistiques F de la première étape) se retrouve en annexe. La dernière ligne du tableau 7 présente les estimations du modèle probit bivarié dans lequel les variables de l'ensemble C tiennent lieu de restrictions d'exclusion. L'estimation de l'incidence à partir du modèle probit bivarié correspond à l'effet marginal estimé en fonction de la moyenne des covariables dans l'équation des résultats.

Avant de passer aux estimations elles-mêmes, nous examinons brièvement la condition de support pour les estimations d'appariement. Les problèmes de support se manifestent lorsqu'il y a des intervalles où les scores de propension affichent une densité positive pour les participants, mais une densité nulle ou très faible pour les non-participants. Le problème de support signifie qu'on ne dispose d'aucun non-participant (ou de très peu) pour construire le contre-factuel estimé des participants. Nous n'avons pas éprouvé de problème de support dans le cas des estimations d'appariement pour l'ensemble des types de formation et pour la formation financée par l'employeur, tant chez les hommes que chez les femmes. Des problèmes mineurs de support se produisent à certaines valeurs supérieures de probabilité de participation dans le cas de la formation financée par l'État et de la formation autofinancée. Étant donné que la proportion de participants dans ces groupes est faible, le problème de support a vraisemblablement une importance secondaire par rapport aux autres questions examinées ci-dessous.

Les effets estimés de tous les types de formation figurant au tableau 7 révèlent plusieurs tendances importantes. Premièrement, les estimations obtenues à partir du modèle de probabilité linéaire (MPL), du modèle probit (converties sous la forme d'effets marginaux pour en assurer la comparabilité) et de l'appariement fondé sur les variables de l'ensemble A (soit les variables utilisées dans le MPL et le modèle probit) sont relativement similaires. Cela indique que la restriction de la forme fonctionnelle linéaire implicite dans le MPL et le modèle probit n'est pas déterminante ici. Deuxièmement, l'appariement fondé sur les variables de l'ensemble B, soit les variables relatives à la situation au chapitre de l'emploi et aux caractéristiques de l'emploi en 1997, réduit les estimations de l'incidence de plus de 50 %. Ces résultats illustrent l'intérêt que présente le conditionnement sur les variables

« endogènes » dans le processus d'appariement. Manifestement, les variables de l'ensemble B permettent de cerner des différences entre les participants et les non-participants que les variables de l'ensemble A ne parviennent pas à dégager.

Troisièmement, les estimations de l'incidence tirées des VI et du modèle probit bivarié sont énormes — bien trop élevées pour être plausibles compte tenu de l'importance du traitement évalué, dont la durée moyenne s'établit autour de 320 heures selon le tableau 3. Nous examinons la vraisemblance de toutes les estimations de manière plus poussée dans notre analyse des gains ci-dessous. Les valeurs aberrantes des estimations fondées sur les VI et le modèle normal bivarié indiquent que l'ensemble C, qui d'un point de vue théorique contient les facteurs les plus prometteurs pour tenir lieu d'instruments valides dans les données de l'EEFA, ne renferme pas en fait d'instruments valides. Pour vérifier cette observation, nous avons répété l'analyse en utilisant chacune des variables de l'ensemble C comme instrument unique à tour de rôle; cet exercice n'a pas changé les résultats qualitatifs obtenus. Nous avons également soumis les instruments à des tests de « Hausman ». La procédure de base dans ce cas consiste à présumer que tous les instruments éventuels sauf un sont valides et à tester ce dernier conditionnellement aux autres. On répète ensuite cette procédure pour chacun des instruments éventuels (ou pour chaque ensemble d'instruments dans le cas de variables dichotomiques représentant des catégories d'une variable commune). Étonnamment, nous n'avons pu rejeter l'hypothèse nulle des instruments valides que dans un cas. La faiblesse de ce test, bien entendu, est qu'il importe qu'au moins certains des instruments soient valides et, étant donné les estimations obtenues, il n'est pas clair que nous répondions à cette exigence.

Quatrièmement, les estimations visant les hommes et les femmes sont relativement similaires.

Les tableaux 8A et 8B présentent des ensembles analogues d'estimations de l'incidence pour chacun des trois types de formation : financée par l'employeur, autofinancée et financée par l'État[8]. Les estimateurs s'apparentent à ceux du tableau 7, à quelques exceptions près. Pour estimer les effets des trois types de formation dans le modèle de BCG, nous avons simplement inclus des indicateurs pour les trois types de formation dans le MPL et le modèle probit au lieu de l'indicateur unique pour tous les types de formation au tableau 7. Dans le cas des estimations d'appariement, nous avons produit des estimations distinctes de l'incidence pour chaque type de formation en utilisant les participants au type de formation en question comme groupe de traitement et les individus n'ayant participé à aucune formation comme groupe de référence. Nous appliquons aux estimations des VI la stratégie employée dans le cas des estimations de BCG, c'est-à-dire que nous remplaçons l'indicateur unique de formation par des indicateurs distincts pour chacun des trois types de formation. Nous pouvons procéder ainsi parce que l'ensemble C contient plus de trois variables. En effet, comme on le sait bien, pour remplir la condition de rang, on doit avoir au moins un instrument pour chaque variable endogène. Enfin, nous ne présentons aucune estimation selon le type de formation pour le modèle normal bivarié. La généralisation de ce modèle à de multiples types de formation nécessite une estimation conjointe d'une équation probit des résultats et d'un modèle probit multinomial pour une participation à la formation associée à quatre options. Il s'agit là d'une tâche qui déborde le cadre du présent rapport.

N° 81-595-MIF2003008 au Catalogue

Passons maintenant aux estimations selon la source de financement : les résultats sont encore plus intéressants. Le résultat le plus évident dans les tableaux 8A et 8B est que les estimations pour tous les types de formation dans le tableau 7 cachent un niveau élevé d'hétérogénéité selon le type en ce qui a trait à l'incidence de la formation. Les données des tableaux 8A et 8B indiquent que la formation financée par l'employeur a une incidence positive, assez considérable d'après la plupart des indicateurs, que la formation financée par l'État a une incidence négative de faible à importante, tandis que la formation autofinancée a une incidence positive ou négative selon l'estimateur examiné dans le cas des hommes et une incidence faible mais systématiquement positive dans le cas des femmes.

La plupart des autres tendances relevées au tableau 7 se retrouvent dans les tableaux 8A et 8B. Les estimations d'appariement fondées sur les variables de l'ensemble A restent très semblables à celles des modèles de probabilité linéaire et des modèles probits axés sur les mêmes covariables. L'ajout des variables de l'ensemble B réduit, encore une fois, l'importance des estimations de l'incidence — de façon encore plus marquée ici que dans le tableau 7. Dans le cas des hommes, par exemple, les estimations de l'incidence de la formation financée par l'employeur baissent d'environ 12 points de pourcentage à environ 1 point de pourcentage lorsque l'appariement porte sur les variables des ensembles A et B plutôt que sur celles de l'ensemble A seulement. Là aussi, l'avantage théorique de l'appariement, qui permet des comparaisons conditionnelles à des variables qui seraient endogènes dans un modèle de régression, a des conséquences pratiques importantes. Comme dans le tableau 7, l'estimateur des VI nous donne des estimations invraisemblables que nous attribuons, ici aussi, à la médiocrité des instruments disponibles dans le cadre de l'EEFA. Les tendances générales des estimations restent similaires pour les hommes et les femmes.

Sur le fond, seules les estimations d'appariement axées sur les ensembles de variables A et B semblent raisonnables. Les autres estimations sont toutes trop élevées pour être plausibles, compte tenu de la durée moyenne de la formation dont nous cherchons à mesurer l'incidence. Même l'appariement axé sur les variables des ensembles A et B produit des estimations d'incidence négatives pour la formation financée par l'État. Même si on a déjà relevé certains effets négatifs associés à la formation financée par les administrations publiques — notamment chez les jeunes hommes dans la *U.S. National Job Training Partnership Act Study* (NJS) dont font état Bloom et al. (1993), nous hésitons à tirer des conclusions de cette nature pour la formation financée par l'État, surtout parce que la NJS a produit des effets positifs chez les adultes, groupe dont l'âge correspond davantage à la population à l'étude dans le présent rapport.

4.2 Gains hebdomadaires habituels

Le tableau 9 (tous les types de formation) et les tableaux 10A et 10B (formation selon le type) présentent l'incidence estimée sur les gains hebdomadaires habituels[9]. Ces tableaux reprennent la structure fondamentale des tableaux 7, 8A et 8B pour ce qui est des estimateurs présentés, à quelques différences mineures près qui s'expliquent par le passage d'une variable de résultats binaire à une variable continue. Premièrement, au lieu de présenter un MPL et un modèle probit pour le modèle de BCG, nous proposons une seule régression linéaire. Deuxièmement, l'estimateur

normal bivarié combine maintenant une équation probit de participation et une équation linéaire de résultats; les estimations présentées sont issues de l'estimation (conjointe) de ce modèle axée sur l'ensemble de l'information plutôt que de la méthode traditionnelle en deux étapes de Heckman (1979). Compte tenu de la disponibilité de cette version plus efficace (c.-à-d. donnant lieu à des écarts-types plus petits) de ce modèle fondée sur l'information complète dans les progiciels courants, il n'y a pas de raison de recourir à la méthode en deux étapes. Comme c'est le cas dans la section précédente, nous ne présentons pas d'estimations du modèle normal bivarié pour chacun des types de formation, l'extension du modèle normal aux traitements multiples, telle que la décrivent Lee (1983) ou Dubin et McFadden (1984), débordant le cadre de ce rapport.

Il n'est pas surprenant de constater que l'incidence de tous les types de formation sur les gains apparaissant au tableau 9 suit à peu près la courbe observée dans le cas de l'incidence sur l'emploi. Relevons notamment la similarité entre les estimations de régression linéaire et les estimations d'appariement fondées sur les variables de l'ensemble A, la forte baisse de la valeur des estimations lorsqu'on passe d'un appariement portant sur les variables de l'ensemble A à un appariement portant sur les variables des ensembles A et B, l'invraisemblance incontestable des estimations des VI et du modèle normal bivarié et la similarité générale des estimations chez les hommes et les femmes. Ces mêmes tendances se retrouvent aussi lorsque nous décomposons la formation selon la source de financement dans les tableaux 10A et 10B. Nous observons encore une fois des écarts importants entre les estimations selon la source de financement, une forte incidence positive sur les gains étant associée à la formation financée par l'employeur et une forte incidence négative étant généralement associée à la formation financée par l'État. La seule différence par rapport à l'incidence sur l'emploi est que l'incidence sur les gains dans le cas de la formation autofinancée est généralement presque aussi négative que celle de la formation financée par l'État.

4.3 Vraisemblance des estimations de l'incidence

Dans un contexte où les gains sont la variable dépendante, il est relativement plus facile d'évaluer la vraisemblance des effets estimés. Examinons deux points de repère. Le premier se retrouve dans les ouvrages traitant de l'incidence d'une année de scolarité sur les gains. Card (1999) propose un résumé récent de ces écrits. La valeur type attribuée à l'incidence d'une année de scolarité sur les gains, établie à partir d'une équation logarithmique des salaires où l'on contrôle dans une certaine mesure le biais dû aux capacités, s'élève à environ 8 %. Examinons maintenant les gains moyens des répondants à l'EEFA. Comme le montre le tableau 4, la moyenne globale pour les hommes se chiffre à environ 500 $ par semaine, et celle des femmes, à environ 300 $ par semaine. La durée moyenne d'une année de scolarité est d'environ 32 semaines; la durée moyenne de la formation dans les données de l'EEFA (tableau 3) est d'environ huit semaines (320 heures / 40 heures par semaine). Par conséquent, si la formation mesurée dans l'EEFA se caractérise par un taux d'efficacité horaire comparable à celui des études scolaires, elle devrait produire des effets sur les gains d'environ 2 % (= 8/32 * 0,08), ce qui correspond à un montant d'environ 10 $ par semaine pour les hommes et d'environ 6 $ par semaine pour les femmes.

N° 81-595-MIF2003008 au catalogue

Le second point de repère est issu du calcul d'un taux de rendement. Le coût d'opportunité constitue vraisemblablement le principal coût de la formation. Pour les hommes, le coût d'opportunité de huit semaines de formation s'établit à 4 000 $ (= 500 $ par semaine * huit semaines), tandis que pour les femmes, il se chiffre à 2 400 $ environ (= 300 $ par semaine * huit semaines). Posons l'hypothèse d'un taux de rendement annuel de 5 % et que l'incidence de la formation sur les gains persiste indéfiniment. L'investissement de 4 000 $ devrait alors produire une incidence de quelque 200 $ par année ou environ 4 $ par semaine. De même, l'investissement de 2 400 $ devrait donner lieu à une incidence de quelque 120 $ par année, soit environ 2,50 $ par semaine.

De toute évidence, ces deux points de repère restent assez grossiers. Comme l'indique Card (1999), l'incidence d'une année de scolarité sur les gains reste fort controversée. Le calcul du taux brut de rendement ne tient pas compte des écarts au chapitre des coûts d'opportunité pour les personnes suivant différents types de formation (comme en témoignent clairement les écarts observés dans les gains hebdomadaires moyens au tableau 4) ni du fait que les personnes tendront à suivre une formation lorsque les coûts d'opportunité sont temporairement faibles, soit en raison d'une demande faible au sein de l'entreprise, soit en raison d'un épisode de chômage. Malgré ces mises en garde, les deux points de repère mettent clairement en évidence l'invraisemblance des estimations excédant 10 $ par semaine.

Les estimations les plus plausibles dans les tableaux 9, 10A et 10B sont celles axées sur l'appariement en fonction des variables des ensembles A et B. La comparaison de ces estimations aux points de repère révèle toutefois que même le niveau de ces estimations est, à quelques exceptions près, beaucoup trop élevé pour être plausible. Par exemple, au tableau 9, ces estimations varient de 33 $ à 45 $ pour les hommes et de 20 $ à 28 $ pour les femmes, selon le nombre de voisins utilisés dans l'appariement par plus proche voisin. Ce sont là des estimations de cinq à dix fois supérieures à ce qui semble raisonnable selon nos points de repère.

4.4 Sommaire des estimations de l'incidence

Nous pouvons résumer nos résultats sous deux angles : la variation des estimations selon différents estimateurs d'analyse économétrique ou le sens fondamental et la vraisemblance des estimations. Dans le premier cas, nos résultats mettent en évidence l'absence d'instruments valides dans les données de l'EEFA, comme en témoignent les estimations extraordinairement invraisemblables obtenues à partir de l'estimateur des VI et de l'estimateur normal bivarié. Conditionnellement aux covariables de l'ensemble A, nous observons peu d'écarts entre les estimations axées sur l'estimateur linéaire type et sur l'appariement semi-paramétrique. L'avantage que présente l'appariement se manifeste lorsqu'on utilise les variables de l'ensemble B, variables qui ne pouvaient être incluses dans les modèles linéaires en raison de considérations d'endogénéité. L'inclusion des variables de l'ensemble B permet de réduire considérablement l'importance des estimations dans tous les cas.

Sur le fond, seules les estimations par appariement axées sur les variables des ensembles A et B combinés semblent à peu près plausibles, compte tenu des résultats de travaux traitant de l'incidence d'une année de scolarité sur les gains, ou si l'on considère les taux normaux de rendement d'un investissement. Mais même là, ces estimations restent nettement supérieures à ce qu'elles devraient être à la lumière de

ces points de repères externes. Le cas de la formation financée par l'État, qui donne lieu à des estimations ponctuelles négatives et statistiquement significatives, fait exception à cet égard. Ces résultats sont trop négatifs même par rapport aux normes peu élevées établies dans les ouvrages publiés.

Par conséquent, notre conclusion générale tirée de ces travaux empiriques est qu'il importe d'apporter certaines modifications à l'EEFA si on entend l'employer comme outil d'estimation des effets de la formation en plus de son rôle principal d'outil de mesure de la fréquence et de la nature de l'éducation et de la formation des adultes. Dans un article connexe, Hui et Smith (2003), nous analysons en détail les modifications proposées à l'enquête qui pourraient faire de l'EEFA un outil d'évaluation plus efficace. Sans ces changements, il convient, à la lumière de nos résultats, de faire preuve d'une très grande prudence dans la production et l'interprétation d'estimations de l'incidence fondées sur ces données. Dans la section qui suit, nous résumons brièvement les principaux thèmes abordés dans Hui et Smith (2003).

N° 81-595-MIF2003008 au catalogue

5. Problèmes relatifs à l'EEFA aux fins de l'évaluation de l'incidence

Hui et Smith (2003) présentent une liste détaillée de suggestions visant à accroître l'utilité de l'EEFA aux fins de l'estimation des effets de l'éducation et de la formation des adultes sur la situation sur le marché du travail au Canada. Nous en résumons ici les quatre principaux thèmes.

Le premier thème abordé dans Hui et Smith (2003) est d'ordre général. De meilleures données sur le moment où sont entreprises des activités de formation amélioreraient substantiellement l'EEFA. À l'heure actuelle, il est impossible de déterminer à quel moment prend fin un épisode de formation par rapport à la date de l'enquête, c'est-à-dire au moment où les résultats sont mesurés. Dans la structure actuelle de l'enquête, cet intervalle peut varier d'un mois à 14 mois, ce qui a des conséquences tant sur l'interprétation des mesures de l'incidence obtenues que sur les tentatives de mesurer la progression temporelle de l'incidence. Une considération connexe se rapporte au moment des épisodes d'emploi pendant l'année précédant l'enquête. De meilleurs renseignements à ce chapitre permettraient aux chercheurs de coupler systématiquement les épisodes de formation à des emplois précis.

Les trois autres thèmes portent tous sur la production de données adaptées à des méthodes particulières d'analyse économétrique. Ainsi, le deuxième thème abordé dans Hui et Smith (2003) concerne l'intérêt de recueillir des données sur des covariables supplémentaires qui seront utilisées de pair avec des estimateurs supposant une sélection en fonction des facteurs observables. Il pourrait s'agir de mesures du contexte familial (niveau de scolarité des parents par exemple), de mesures plus détaillées de l'obtention de diplômes ou de certificats, ou encore d'une mesure des capacités. Le troisième thème se rapporte à l'intérêt d'obtenir des mesures répétées des résultats (ou à tout le moins une mesure des résultats) antérieures à la formation afin de permettre l'application de méthodes d'évaluation longitudinales. Enfin, le quatrième thème concerne l'intérêt de disposer d'instruments plus crédibles associés aux données de l'EEFA. Il pourrait s'agir de renseignements recueillis auprès des répondants sur le centre de formation le plus proche ou sur les subventions à la formation offertes par l'entreprise, ou encore de données couplées provenant de sources externes pour des variables telles que les droits de scolarité dans les collèges publics de chacune des provinces. L'intérêt que présente chacune de ces avenues d'amélioration de l'EEFA se dégage clairement de l'analyse rapportée dans cet article.

6. Conclusions

Nos résultats inspirent quelques conclusions importantes. Premièrement, les estimations de l'incidence axées sur l'EEFA se caractérisent par une sensibilité extrême au choix de l'estimateur non expérimental de l'incidence. Il est essentiel d'examiner avec rigueur la vraisemblance des méthodes d'analyse économétrique lorsqu'on utilise ces données pour estimer l'incidence de l'éducation et de la formation des adultes.

Deuxièmement, nos résultats mettent en évidence l'absence d'un instrument plausible pour la formation parmi les variables disponibles dans l'EEFA de 1998. Par conséquent, nous obtenons des estimations d'incidence extrêmement élevées lorsque nous utilisons l'estimateur des VI et l'estimateur normal bivarié — tellement élevées qu'elles en sont invraisemblables.

Troisièmement, nous observons peu d'écart entre les estimations obtenues à partir des méthodes d'appariement semi-paramétriques et les modèles linéaires traditionnels, comme la régression ou le modèle probit lorsqu'on utilise un ensemble commun de covariables. L'avantage de l'appariement dans cette application est qu'il permet le conditionnement sur des variables relatives à la situation d'activité et aux caractéristiques de l'emploi au cours de l'année précédente. L'inclusion de telles variables dans un modèle linéaire poserait des problèmes d'endogénéité. Le recours à ces variables — appelées variables de l'ensemble B dans notre analyse — permet de réduire considérablement l'importance des effets estimés. En fait, les seules estimations vraiment plausibles sont celles fondées sur l'appariement en fonction des variables des ensembles A et B combinés. Ces résultats soulèvent deux points méthodologiques déjà abordés dans les écrits mais dont on fait souvent abstraction : d'une part, les variables sur lesquelles porte l'appariement conditionnent dans une large mesure les résultats obtenus et, d'autre part, l'appariement peut produire des résultats très médiocres s'il s'applique à un ensemble de covariables ne respectant pas l'hypothèse d'indépendance conditionnelle qui justifie l'appariement.

Quatrièmement, même les estimations les plus plausibles — celles obtenues par l'appariement en fonction des variables des ensembles A et B, ce dernier comprenant des variables mesurant la situation d'activité et les caractéristiques de l'emploi au cours de l'année précédente — sont invraisemblables. Lorsqu'on les compare aux points de repères relatifs au rendement estimé d'une année de scolarité ou au simple rendement d'un investissement, ces estimations semblent être au moins de cinq à dix fois trop élevées.

Enfin, la conclusion générale de notre étude est la suivante : l'EEFA, sous sa forme actuelle, ne constitue pas un très bon outil pour l'évaluation de l'incidence de l'éducation et de la formation des adultes. Cela n'est peut-être pas surprenant dans la mesure où cette enquête a été conçue pour mesurer la fréquence et la diversité de l'éducation et de la formation des adultes. Il importe néanmoins de le souligner.

N° 81-595-MIF2003008 au catalogue

Selon nous, toutefois, l'enquête pourrait servir à cette fin si on apportait les modifications proposées dans Hui et Smith (2003). La constitution d'un ensemble de données, qu'il s'agisse des données provenant de l'EEFA ou d'autres sources, que l'on pourrait aisément utiliser pour évaluer l'incidence de l'éducation et de la formation des adultes serait bénéfique tant du point de vue des politiques que de la recherche au Canada.

Tableaux

Tableau 1

Total d'observations totales et restrictions sur l'échantillon de l'EEFA de 1998

	Hommes	Femmes
Nombre total d'observations	14 875	18 535
Étudiants à temps plein en 1998	855	1 107
Non-étudiants à temps plein en 1998		
Âge : 17 à 24 ans	856	1 052
Âge : 65 ans et plus	2 416	3 958
Échantillon d'analyse	10 748	12 418

Notes : Les statistiques présentées sont pondérées à l'aide des facteurs de pondération fournis par Statistique Canada pour tenir compte de l'échantillon stratifié et de la non-réponse.

Tableau 2

Statistiques sommaires sur l'incidence de la formation

	Hommes		Femmes	
	Pourcentage			
Participation à un programme				
Programmes financés par l'employeur	3,08	(0,23)	2,07	(0,17)
Programmes financés par l'État	0,84	(0,14)	1,21	(0,15)
Programmes autofinancés	2,87	(0,25)	4,61	(0,29)
Tous les programmes de formation	6,83	(0,36)	7,91	(0,36)
Participation à un cours				
Cours financés par l'employeur	17,07	(0,51)	15,52	(0,47)
Cours financés par l'État	1,31	(0,19)	1,28	(0,14)
Cours autofinancés	3,79	(0,26)	6,78	(0,35)
Tous les cours de formation	21,90	(0,57)	22,93	(0,55)
Participation à la formation (programmes et cours)				
Formation financée par l'employeur	19,39	(0,53)	17,00	(0,48)
Formation financée par l'État	2,11	(0,23)	2,40	(0,20)
Formation autofinancée	6,38	(0,35)	10,81	(0,43)
Tous les genres de formation	27,07	(0,62)	28,76	(0,60)

Notes : Les écarts-types estimés figurent entre parenthèses. La taille des échantillons est de 10 748 et 12 418 pour les hommes et les femmes respectivement.

Les statistiques présentées sont pondérées à l'aide des facteurs de pondération fournis par Statistique Canada pour tenir compte de l'échantillon stratifié et de la non-réponse.

N° 81-595-MIF2003008 au catalogue

Tableau 3

Statistiques sommaires sur la durée totale de la formation exprimée en heures

	Moyenne	Quartiles 1er quartile	Médiane	3e quartile
Tous les types de formation				
Hommes				
Formation financée par l'employeur	69,1	12	24	48
Formation financée par l'État	591,5	90	425	1 026
Formation autofinancée	526,2	32	304	960
Tous les types de formation	320,7	16	48	425
Femmes				
Formation financée par l'employeur	57,7	8	18	42
Formation financée par l'État	605,3	99	480	1 000
Formation autofinancée	443,4	24	180	745
Tous les types de formation	314,7	12	50	450
Programmes de formation				
Hommes				
Formation financée par l'employeur	252,5	40	90	240
Formation financée par l'État	770,4	320	736	1 170
Formation autofinancée	730,6	225	600	1 120
Tous les types de formation	676,1	156	480	1 080
Femmes				
Formation financée par l'employeur	219,2	39	96	280
Formation financée par l'État	783,5	352	720	1 200
Formation autofinancée	646,4	180	496	990
Tous les types de formation	624,6	150	480	960
Cours de formation				
Hommes				
Formation financée par l'employeur	34,9	10	18	38
Formation financée par l'État	75,3	6	19	50
Formation autofinancée	66,9	6	18	40
Tous les types de formation	45,6	7	18	42
Femmes				
Formation financée par l'employeur	34,7	6	18	30
Formation financée par l'État	117,8	10	30	112
Formation autofinancée	62,7	6	16	35
Tous les types de formation	50,6	6	18	36

Notes : Les statistiques utilisent les bornes inférieures pour les épisodes de formation en cours au moment de l'entrevue de l'EEFA et pour les épisodes tronqués par le haut. Voir la discussion dans le rapport à ce sujet. L'échantillon comprend les répondants, définis au tableau 1, ayant déclaré un nombre positif d'heures de formation.

Les statistiques présentées sont pondérées à l'aide des facteurs de pondération fournis par Statistique Canada pour tenir compte de l'échantillon stratifié et de la non-réponse.

Tableau 4

Statistiques sommaires – Variables de résultats employées dans l'analyse

	Gains hebdomadaires moyens ($)				Emploi (%)			
	Hommes		Femmes		Hommes		Femmes	
Total	477,54	(6,63)	302,84	(4,73)	80,01	(0,55)	67,32	(0,60)
Aucune formation	413,91	(7,38)	236,25	(5,16)	76,07	(0,68)	60,00	(0,74)
Tous les types de formation	648,96	(13,28)	467,83	(9,35)	90,62	(0,82)	85,46	(0,84)
Formation financée par l'employeur	748,02	(14,68)	582,70	(10,35)	96,31	(0,49)	96,58	(0,52)
Formation financée par l'État	331,63	(51,10)	186,13	(23,74)	61,26	(5,52)	48,84	(4,23)
Formation autofinancée	448,53	(23,39)	367,75	(16,81)	83,44	(2,10)	77,36	(1,62)

Notes : Les écarts-types estimés apparaissent entre parenthèses. Les résultats sont mesurés en mars 1998 pour les répondants du Québec et en janvier 1998 pour tous les autres répondants. La formation renvoie à la formation reçue au cours de l'année civile 1997.

Les statistiques présentées sont pondérées à l'aide des facteurs de pondération fournis par Statistique Canada pour tenir compte de l'échantillon stratifié et de la non-réponse.

Tableau 5

Caractéristiques des répondants et de l'emploi

	Hommes		Femmes	
Province				
Terre-Neuve	1,91	(0,10)	1,89	(0,09)
Île-du-Prince-Édouard	0,44	(0,03)	0,44	(0,02)
Nouvelle-Écosse	2,99	(0,13)	3,13	(0,13)
Nouveau-Brunswick	2,51	(0,11)	2,53	(0,10)
Québec	24,91	(0,62)	24,92	(0,58)
Ontario	38,12	(0,72)	38,29	(0,68)
Manitoba	3,51	(0,15)	3,47	(0,14)
Saskatchewan	2,96	(0,13)	2,97	(0,12)
Alberta	9,49	(0,36)	9,20	(0,33)
Colombie-Britannique	13,16	(0,46)	13,16	(0,43)
Région				
Région métropolitaine de recensement	66,01	(0,58)	64,96	(0,56)
Toronto/Vancouver/Montréal	35,05	(0,77)	35,54	(0,72)
Centre urbain	7,25	(0,27)	7,98	(0,29)
Région rurale	16,07	(0,40)	15,96	(0,38)
Groupe d'âge				
25 à 34	27,55	(0,64)	27,22	(0,59)
35 à 44	31,52	(0,63)	31,50	(0,60)
45 à 54	24,92	(0,60)	24,87	(0,58)
55 à 64	16,01	(0,50)	16,41	(0,49)
Âge (années)	42,40	(0,15)	42,50	(0,14)
Âge au carré	1 907,22	(12,82)	1 917,11	(12,23)
Niveau de scolarité				
Études secondaires partielles	13,82	(0,48)	13,00	(0,40)
11e - 13e année terminée	18,79	(0,56)	21,75	(0,56)
Études postsecondaires partielles	7,38	(0,36)	8,19	(0,35)
Certificat ou diplôme postsecondaire	32,49	(0,63)	32,75	(0,61)
Baccalauréat, maîtrise ou doctorat	20,11	(0,58)	17,10	(0,51)
Présence d'un conjoint	72,09	(0,64)	71,68	(0,57)
Niveau de scolarité du conjoint				
Études secondaires partielles	9,25	(0,39)	9,70	(0,36)
11e - 13e année terminée	16,73	(0,52)	13,45	(0,46)
Études postsecondaires partielles	6,05	(0,33)	4,76	(0,27)
Certificat ou diplôme postsecondaire	23,93	(0,58)	24,17	(0,56)
Baccalauréat, maîtrise ou doctorat	39,84	(0,69)	42,18	(0,65)
Nombre d'enfants (âgés de moins de 18 ans)				
1 enfant	17,83	(0,53)	19,37	(0,53)
2 enfants et plus	31,87	(0,66)	34,95	(0,63)
Nombre d'enfants d'âge préscolaire				
1 enfant d'âge préscolaire	11,86	(0,43)	13,24	(0,43)
2 enfants et plus d'âge préscolaire	6,00	(0,32)	6,22	(0,31)
Situation en regard des conjoints et des enfants				
Présence d'un conjoint et d'enfants	47,59	(0,70)	45,41	(0,66)
Présence d'un conjoint et de 2 enfants et plus	31,10	(0,65)	30,56	(0,62)
Présence d'un conjoint et d'enfants d'âge préscolaire	17,64	(0,52)	17,16	(0,49)
Autres caractéristiques du répondant				
Né à l'étranger	19,68	(0,65)	20,55	(0,64)

Tableau 5

Caractéristiques des répondants et de l'emploi (fin)

	Hommes		Femmes	
Caractéristiques de l'emploi				
Employé en 1997	87,11	(0,47)	73,50	(0,57)
Travail à temps plein en 1997	72,85	(0,61)	45,66	(0,66)
Emploi professionnel, administratif ou de direction	28,30	(0,63)	30,41	(0,61)
Emploi de col bleu	38,43	(0,67)	8,35	(0,42)
Syndiqué	26,35	(0,60)	21,63	(0,54)
Taille de l'entreprise : moins de 20 employés	29,43	(0,62)	23,59	(0,54)
20 à 99 employés	13,00	(0,49)	9,93	(0,39)
100 à 199 employés	5,59	(0,32)	4,41	(0,28)
200 à 499 employés	6,21	(0,34)	5,91	(0,32)
500 employés et plus	31,18	(0,65)	27,71	(0,60)
Indicateurs manquants				
Lieu de naissance	0,52	(0,12)	0,37	(0,09)
Affiliation syndicale	0,11	(0,03)	0,07	(0,02)
Taille de l'entreprise	1,58	(0,18)	1,81	(0,19)
Profession	0,46	(0,13)	0,08	(0,03)

Notes : Les écarts-types estimés apparaissent entre parenthèses. Les unités utilisées sont les pourcentages pour les variables binaires et les moyennes pour les variables continues.

Les statistiques présentées sont pondérées à l'aide des facteurs de pondération fournis par Statistique Canada pour tenir compte de l'échantillon stratifié et de la non-réponse.

Tableau 6

Définitions des instruments et des ensembles de conditionnement

Liste des variables de l'ensemble A	
Variables dichotomiques des provinces et des régions de résidence	
Terre-Neuve	Île-du-Prince-Édouard
Nouvelle-Écosse	Nouveau-Brunswick
Québec	Manitoba
Saskatchewan	Alberta
Colombie-Britannique	
Région métropolitaine de recensement	Toronto/Montréal/Vancouver
Centre urbain	Région rurale
Caractéristiques des répondants	
Âge	Âge au carré
Études secondaires partielles	11e - 13e année terminée
Études postsecondaires partielles	Certificat ou diplôme postsecondaire
Baccalauréat/maîtrise/doctorat	
Né à l'étranger	Lieu de naissance manquant
Liste des variables de l'ensemble B	
Caractéristiques de l'emploi en 1997	
Situation d'activité en 1997	Travail à temps plein en 1997
Syndiqué	Emploi de col bleu
Taille de l'entreprise : 20 à 99 employés	Emploi professionnel, administratif
Taille de l'entreprise : 200 à 499 employés	ou de direction
Taille de l'entreprise : 100 à 199 employés	
Taille de l'entreprise : 500 employés et plus	Données manquantes sur l'affiliation syndicale
Données manquantes sur la profession	Données manquantes sur la taille de l'entreprise
Liste des variables de l'ensemble C	
Caractéristiques familiales	
Présence d'un conjoint	
Conjoint : Études secondaires partielles	Conjoint : 11e - 13e année terminée
Conjoint : Études postsecondaires partielles	Conjoint : Certificat ou diplôme
Conjoint : Baccalauréat/maîtrise/doctorat	postsecondaire
1 enfant	2 enfants et plus
1 enfant et un conjoint	2 enfants et plus et un conjoint
1 enfant d'âge préscolaire	2 enfants et plus d'âge préscolaire
Enfants d'âge préscolaire et un conjoint	

No 81-595-MIF2003008 au catalogue

Tableau 7

Incidence estimée de tous les types de formation sur l'emploi

	Incidence sur l'emploi (%)					
	Hommes			Femmes		
	Tous les types de formation			Tous les types de formation		
Modèle de probabilité linéaire	7,82	(1,04)	***	15,91	(1,23)	***
Modèle probit : effet marginal	9,15	(1,18)	***	18,60	(1,38)	***
Appariement par plus proche voisin (ensemble A)						
1 à 1	7,89	(0,97)	***	17,06	(1,08)	***
1 à 2	7,72	(0,81)	***	16,77	(0,91)	***
1 à 5	9,01	(0,71)	***	16,63	(0,78)	***
Appariement par plus proche voisin (ensembles A et B)						
1 à 1	2,66	(0,89)	***	2,59	(0,96)	***
1 à 2	2,99	(0,78)	***	2,78	(0,87)	***
1 à 5	2,49	(0,70)	***	2,87	(0,80)	***
Variable instrumentale	65,56	(12,74)	***	163,06	(34,59)	***
Probit bivarié : effet marginal	66,53	(0,29)	***	69,04	(0,23)	

* Coefficients statistiquement significatifs à 10 %.

** Coefficients statistiquement significatifs à 5 %.

*** Coefficients statistiquement significatifs à 1 %.

Notes : Les écarts-types estimés (robustes) apparaissent entre parenthèses. Les estimations de l'incidence portent sur la formation reçue au cours de l'année civile 1997. Les estimations ne tiennent pas compte des répondants dont la formation était en cours au moment de l'entrevue de l'EEFA. Les échantillons comptent 10 469 hommes et 11 938 femmes. L'emploi est mesuré en mars 1998 pour les répondants du Québec et en janvier 1998 pour tous les autres répondants. Le modèle de probabilité linéaire est fondé sur les ensembles de covariables A et C. Les scores de propension pour l'estimateur d'appariement par plus proche voisin sont fondés sur l'ensemble de covariables A ou sur les ensembles de covariables A et B. Le modèle des variables instrumentales et le modèle probit bivarié appliquent le conditionnement à l'ensemble de covariables A et utilisent les covariables de l'ensemble C comme instruments.

Les coefficients présentés sont pondérés à l'aide des facteurs de pondération fournis par Statistique Canada pour tenir compte de l'échantillon stratifié et de la non-réponse.

Tableau 8A

Incidence estimée de la formation sur l'emploi selon le type de formation – Hommes

	Incidence sur l'emploi (%)					
	Financée par l'employeur		Financée par l'État		Autofinancée	
Modèle de probabilité linéaire	12,49	(0,81) ***	-16,14	(5,12) ***	-0,35	(2,48)
Modèle probit : effet marginal	31,28	(1,07) ***	-19,30	(4,60) ***	4,30	(2,24) *
Appariement par plus proche voisin (ensemble A)						
1 à 1	11,49	(0,98) ***	-18,00	(4,38) ***	-3,61	(2,30)
1 à 2	12,15	(0,80) ***	-17,75	(3,98) ***	-3,61	(2,04) *
1 à 5	12,79	(0,67) ***	-14,3	(3,70) ***	-1,45	(1,91)
Appariement par plus proche voisin (ensembles A et B)						
1 à 1	0,94	(0,73)	-5,50	(4,73)	2,61	(2,48)
1 à 2	0,52	(0,62)	-3,75	(4,35)	1,81	(2,11)
1 à 5	1,00	(0,57) *	-1,4	(3,95)	3,17	(1,89) *
Variable instrumentale	70,61	(16,65) ***	-40,98	(141,07)	-132,70	(54,02) **

* Coefficients statistiquement significatifs à 10 %.
** Coefficients statistiquement significatifs à 5 %.
*** Coefficients statistiquement significatifs à 1 %.

Notes : Les écarts-types estimés (robustes) apparaissent entre parenthèses. Les estimations de l'incidence portent sur la formation reçue au cours de l'année civile 1997. Les estimations ne tiennent pas compte des répondants dont la formation était en cours au moment de l'entrevue de l'EEFA. L'effectif de l'échantillon du modèle de probabilité linéaire et des modèles de VI est de 10 469. Les échantillons de l'appariement par plus proche voisin comptent des effectifs de 10 611, 10 720 et 10 631 pour la formation financée par l'employeur, la formation financée par l'État et la formation autofinancée, respectivement. L'emploi est mesuré en mars 1998 pour les répondants du Québec et en janvier 1998 pour tous les autres répondants. Le modèle de probabilité linéaire est fondé sur les ensembles de covariables A et C. Les scores de propension pour l'estimateur d'appariement par plus proche voisin sont fondés sur l'ensemble de covariables A ou sur les ensembles de covariables A et B. Le modèle des variables instrumentales et le modèle probit bivarié appliquent le conditionnement à l'ensemble de covariables A et utilisent les covariables de l'ensemble C comme instruments.

Les coefficients présentés sont pondérés à l'aide des facteurs de pondération fournis par Statistique Canada pour tenir compte de l'échantillon stratifié et de la non-réponse.

Tableau 8B

Incidence estimée de la formation sur l'emploi selon le type de formation – Femmes

	Incidence sur l'emploi (%)					
	Financée par l'employeur		Financée par l'État		Autofinancée	
Modèle de probabilité linéaire	25,24	(0,98) ***	-17,91	(4,33) ***	4,09	(1,96) **
Modèle probit : effet marginal	30,76	(1,17) ***	-20,32	(5,16) ***	3,57	(2,24)
Appariement par plus proche voisin (ensemble A)						
1 à 1	26,29	(1,15) ***	-24,33	(4,10) ***	1,13	(2,00)
1 à 2	25,92	(0,89) ***	-23,95	(3,71) ***	1,65	(1,77)
1 à 5	27,29	(0,67) ***	-22,74	(3,36) ***	1,79	(1,59)
Appariement par plus proche voisin (ensembles A et B)						
1 à 1	2,77	(0,72) ***	-6,84	(4,35)	3,40	(2,02) *
1 à 2	2,90	(0,61) ***	-4,75	(3,93)	3,76	(1,81) **
1 à 5	3,20	(0,53) ***	-6,62	(3,60) *	3,21	(1,61) **
Variable instrumentale	188,60	(64,36) ***	-72,77	(110,33)	156,43	(106,65)

* Coefficients statistiquement significatifs à 10 %.
** Coefficients statistiquement significatifs à 5 %.
*** Coefficients statistiquement significatifs à 1 %.

Notes : Les écarts-types estimés (robustes) apparaissent entre parenthèses. Les estimations de l'incidence portent sur la formation reçue au cours de l'année civile 1997. Les estimations ne tiennent pas compte des répondants dont la formation était en cours au moment de l'entrevue de l'EEFA. L'effectif de l'échantillon du modèle de probabilité linéaire et des modèles de VI est de 11 938. Les échantillons de l'appariement par plus proche voisin comptent des effectifs de 12 255, 12 370 et 12 145 pour la formation financée par l'employeur, la formation financée par l'État et la formation autofinancée, respectivement. L'emploi est mesuré en mars 1998 pour les répondants du Québec et en janvier 1998 pour tous les autres répondants. Le modèle de probabilité linéaire est fondé sur les ensembles de covariables A et C. Les scores de propension pour l'estimateur d'appariement par plus proche voisin sont fondés sur l'ensemble de covariables A ou sur les ensembles de covariables A et B. Le modèle des variables instrumentales et le modèle probit bivarié appliquent le conditionnement à l'ensemble de covariables A et utilisent les covariables de l'ensemble C comme instruments.

Les coefficients présentés sont pondérés à l'aide des facteurs de pondération fournis par Statistique Canada pour tenir compte de l'échantillon stratifié et de la non-réponse.

N° 81-595-MIF2003008 au catalogue

Tableau 9

Incidence estimée de tous les types de formation sur les gains hebdomdaires habituels

	Incidence sur les gains hebdomadaires ($)			
	Hommes		Femmes	
	Tous les types de formation		Tous les types de formation	
Modèle linéaire	172,8	(15,8) ***	158,9	(11,4) ***
Appariement par plus proche voisin (ensemble A)				
1 à 1	172,1	(12,5) ***	154,8	(8,84) ***
1 à 2	161,6	(11,1) ***	153,2	(7,9) ***
1 à 5	162,6	(10,1) ***	151,6	(7,2) ***
Appariement par plus proche voisin (ensembles A et B)				
1 à 1	33,4	(12,9) ***	28,1	(9,1) ***
1 à 2	45,2	(11,5) ***	20,4	(8,5) ***
1 à 5	44,5	(10,6) ***	22,0	(7,9) ***
Variable instrumentale	1 086,5	(188,4) ***	909,5	(205,3) ***
Modèle normal bivarié	407,2	(90,40) ***	204,4	(18,50) ***

* Coefficients statistiquement significatifs à 10 %.
** Coefficients statistiquement significatifs à 5 %.
*** Coefficients statistiquement significatifs à 1 %.

Notes : Les écarts-types estimés (robustes) apparaissent entre parenthèses. Les estimations de l'incidence portent sur la formation reçue au cours de l'année civile 1997. Les estimations ne tiennent pas compte des répondants dont la formation était en cours au moment de l'entrevue de l'EEFA. Les échantillons comptent 10 469 hommes et 11 938 femmes. Les gains hebdomadaires habituels sont mesurés en mars 1998 pour les répondants du Québec et en janvier 1998 pour tous les autres répondants. Le modèle de probabilité linéaire est fondé sur les ensembles de covariables A et C. Les scores de propension pour l'estimateur d'appariement par plus proche voisin sont fondés sur l'ensemble de covariables A ou sur les ensembles de covariables A et B. Le modèle des variables instrumentales et le modèle probit bivarié appliquent le conditionnement à l'ensemble de covariables A et utilisent les covariables de l'ensemble C comme instruments.

Les coefficients présentés sont pondérés à l'aide des facteurs de pondération fournis par Statistique Canada pour tenir compte de l'échantillon stratifié et de la non-réponse.

Tableau 10A

Incidence estimée de la formation sur les gains hebdomadaires habituels selon le type de formation – Hommes

	Incidence sur les gains hebdomadaires ($)					
	Financée par l'employeur		Financée par l'État		Autofinancée	
Modèle linéaire	256,6	(17,1) ***	-122,4	(50,5) **	-60,5	(28,4) **
Appariement par plus proche voisin (ensemble A)						
1 à 1	251,5	(14,4) ***	-106,7	(41,2) ***	-94,9	(27,5) ***
1 à 2	248,3	(12,4) ***	-113,3	(34,9) **	-99,5	(24,2) ***
1 à 5	243,4	(11,2) ***	-104,7	(30,9) ***	-88,0	(21,6) ***
Appariement par plus proche voisin (ensembles A et B)						
1 à 1	59,4	(14,2) ***	-6,3	(39,8)	-50,1	(27,3) *
1 à 2	50,2	(12,9) ***	-20,3	(35,4)	-82,8	(23,8) ***
1 à 5	60,6	(11,8) ***	16,2	(31,7)	-74,3	(21,1) ***
Variable instrumentale	1 162,5	(203,6) ***	720,1	(1 736,9)	-969,6	(646,0)

* Coefficients statistiquement significatifs à 10 %.
** Coefficients statistiquement significatifs à 5 %.
*** Coefficients statistiquement significatifs à 1 %.

Notes : Les écarts-types estimés (robustes) apparaissent entre parenthèses. Les estimations de l'incidence portent sur la formation reçue au cours de l'année civile 1997. Les estimations ne tiennent pas compte des répondants dont la formation était en cours au moment de l'entrevue de l'EEFA. L'effectif de l'échantillon du modèle de probabilité linéaire et des modèles de VI est de 10 469. Les échantillons de l'appariement par plus proche voisin comptent des effectifs de 10 611, 10 720 et 10 631 pour la formation financée par l'employeur, la formation financée par l'État et la formation autofinancée, respectivement. Les gains hebdomadaires habituels sont mesurés en mars 1998 pour les répondants du Québec et en janvier 1998 pour tous les autres répondants. Le modèle de probabilité linéaire est fondé sur les ensembles de covariables A et C. Les scores de propension pour l'estimateur d'appariement par plus proche voisin sont fondés sur l'ensemble de covariables A ou sur les ensembles de covariables A et B. Le modèle des variables instrumentales et le modèle probit bivarié appliquent le conditionnement à l'ensemble de covariables A et utilisent les covariables de l'ensemble C comme instruments.

Les coefficients présentés sont pondérés à l'aide des facteurs de pondération fournis par Statistique Canada pour tenir compte de l'échantillon stratifié et de la non-réponse.

Tableau 10B

Incidence estimée de la formation sur les gains hebdomadaires habituels selon le type de formation – Femmes

	Incidence sur les gains hebdomadaires ($)					
	Financée par l'employeur		Financée par l'État		Autofinancée	
Modèle linéaire	256,2	(11,9) ***	-93,0	(23,3) ***	15,1	(19,2)
Appariement par plus proche voisin (ensemble A)						
1 à 1	261,9	(10,5) ***	-139,4	(25,4) ***	-33,8	(15,8) **
1 à 2	248,7	(9,4) ***	-116,1	(21,3) ***	-17,8	(13,8)
1 à 5	253,3	(8,4) ***	-116,4	(19,3) ***	-17,5	(12,5)
Appariement par plus proche voisin (ensembles A et B)						
1 à 1	40,6	(10,8) ***	-19,5	(23,9)	3,0	(15,3)
1 à 2	47,2	(10,0) ***	-31,4	(22,1)	-4,9	(13,8)
1 à 5	48,4	(9,3) ***	-37,3	(20,0) *	-18,7	(12,7)
Variable instrumentale	1 112,1	(391,0) ***	-666,9	(681,5)	1 035,2	(679,1)

* Coefficients statistiquement significatifs à 10 %.
** Coefficients statistiquement significatifs à 5 %.
*** Coefficients statistiquement significatifs à 1 %.

Notes : Les écarts-types estimés (robustes) apparaissent entre parenthèses. Les estimations de l'incidence portent sur la formation reçue au cours de l'année civile 1997. Les estimations ne tiennent pas compte des répondants dont la formation était en cours au moment de l'entrevue de l'EEFA. L'effectif de l'échantillon du modèle de probabilité linéaire et des modèles de VI est de 11 938. Les échantillons de l'appariement par plus proche voisin comptent des effectifs de 12 255, 12 370 et 12 145 pour la formation financée par l'employeur, la formation financée par l'État et la formation autofinancée, respectivement. Les gains hebdomadaires habituels sont mesurés en mars 1998 pour les répondants du Québec et en janvier 1998 pour tous les autres répondants. Le modèle de probabilité linéaire est fondé sur les ensembles de covariables A et C. Les scores de propension pour l'estimateur d'appariement par plus proche voisin sont fondés sur l'ensemble de covariables A ou sur les ensembles de covariables A et B. Le modèle des variables instrumentales et le modèle probit bivarié appliquent le conditionnement à l'ensemble de covariables A et utilisent les covariables de l'ensemble C comme instruments.

Les coefficients présentés sont pondérés à l'aide des facteurs de pondération fournis par Statistique Canada pour tenir compte de l'échantillon stratifié et de la non-réponse.

N° 81-595-MIF2003008 au catalogue

Références

Angrist, Joshua et Alan Krueger. 1999. "Empirical Strategies in Labor Economics." Chapter 23 in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics, Volume 3A*. Amsterdam: Elsevier Science, p. 1277-1366.

Arulampalam, Wiji, Alison Booth et Peter Elias. 1997. "Work-Related Training and Earnings Growth for Young Men in Britain." *Research in Labor Economics*. 16: 119-147.

Barnow, Burt. 1987. "The Impact of CETA Programs on Earnings: A Review of the Literature," *Journal of Human Resources*. 22(2): 157-193.

Barnow, Burt, Glen Cain et Arthur Goldberger. 1980. "Issues in the Analysis of Selectivity Bias." In Ernst Stromsdorfer and George Farkas, eds., *Evaluation Studies Review Annual, Volume 5*. Beverly Hills, CA: Sage Publications, p. 43-59.

Barron, John, Mark Berger et Dan Black. 1997. "How Well Do We Measure Training?" *Journal of Labor Economics*. 15(3, Part 1): 507-528.

Becker, Gary. 1964. *Human Capital*. New York: Columbia University Press for National Bureau of Economic Research.

Björklund, Anders et Robert Moffitt. 1987. "Estimation of Wage Gains and Welfare Gains in Self-Selection Models." *Review of Economics and Statistics*. 69(1): 42-49.

Bloom, Howard, Larry Orr, George Cave, Steven Bell et Fred Doolittle. 1993. *The National JTPA Study: Title II-A Impacts on Earnings and Employment at 18 Months*. Bethesda, MD: Abt Associates for Employment and Training Administration, U.S. Department of Labor.

Bound, John, David Jaeger et Regina Baker. 1995. "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variables is Weak." *Journal of the American Statistical Association*. 90(430): 443-450.

Blundell, Richard, Lorraine Dearden et Costas Meghir. 1996. *The Determinants and Effects of Work-Related Training in Britain*. London: Institute for Fiscal Studies.

Card, David. 1999. "The Causal Effect of Education on Earnings." Chapter 30 in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics, Volume 3A*. Amsterdam: Elsevier Science, p. 1801-1864.

Carneiro, Pedro, James Heckman, et Dayanand Manoli. 2002. "Human Capital Policy." Unpublished manuscript, University of Chicago.

Dehejia, Rajeev et Sadek Wahba. 1999. "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs." *Journal of the American Statistical Association*. 94(448): 1053-1062.

Dubin, Jeffrey et Daniel McFadden. 1984. "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption." *Econometrica*. 52(2): 345-362.

Heckman, James. 1979. "Sample Selection Bias as a Specification Error." *Econometrica*. 47(1): 153-161.

Heckman, James. 1997. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources*. 32(3): 441-461.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith et Petra Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica*. 66(5): 1017-1098.

Heckman, James, Hidehiko Ichimura et Petra Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *Review of Economic Studies*. 64(4): 605-654.

Heckman, James, Hidehiko Ichimura et Petra Todd. 1998. "Matching as an Econometric Evaluation Estimator." *Review of Economic Studies*. 65(2): 261-294.

Heckman, James, Robert LaLonde et Jeffrey Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs." Chapter 31 in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics, Volume 3A*. Amsterdam: Elsevier Science, p. 1866-2097.

Heckman, James, Lance Lochner, Jeffrey Smith et Christopher Taber. 1997. "The Effects of Government Policy on Human Capital Investment and Wage Inequality." *Chicago Policy Review*. 1(2):1-40.

Heckman, James et Richard Robb. 1985. "Alternative Methods for Evaluating the Impact of Interventions." Chapter 4 in James Heckman and Burton Singer, eds., *Longitudinal Analysis of Labor Market Data. Econometric Society Monograph Series, no. 10*. New York: Cambridge University Press, p. 156-245.

Heckman, James et Jeffrey Smith. 1996. "Experimental and Nonexperimental Evaluation." In Günther Schmid, Jacqueline O'Reilly, and Klaus Schömann, eds., *International Handbook of Labour Market Policy and Evaluation*. Brookfield, VT: Edward Elgar. 37-88.

Heckman, James et Jeffrey Smith. 1998. "Evaluating the Welfare State." In Steinar Strom, ed., *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium. Econometric Society Monograph Series, no. 31*. New York: Cambridge University Press, p. 241-318.

Heckman, James, Jeffrey Smith et Nancy Clements. 1997. "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts." *Review of Economic Studies*. 64(4): 487-535.

Heckman, James et Edward Vytlacil. 2001. "Policy-Relevant Treatment Effects." *American Economic Review*. 91(2): 107-111.

Hotz, V. Joseph, Guido Imbens et Jacob Klerman. 2000. "The Long-Term Gains from GAIN: A Re-Analysis of the Impacts of the California GAIN Program." NBER Working Paper No. 8007. Cambridge, MA: National Bureau of Economic Research.

N° 81-595-MIF2003008 au catalogue

Hui, Shek-Wai et Jeffrey Smith. 2003. « Enjeux liés au contenu de l'Enquête canadienne sur l'éducation et sur la formation des adultes » Statistique Canada, nº 81-595-MIF2003009 au Catalogue. Ottawa, Ministre responsable de Statistique Canada.

Imbens, Guido et Joshua Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*. 62(2): 467-475.

Lechner, Michael et Ruth Miquel. 2002. "Identification of Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions." Discussion Paper. Revised March, 2002. St Gallen, Switzerland: Department of Economics, University of St. Gallen.

Lee, Lung-Fei. 1983. "Generalized Econometric Models with Selectivity." *Econometrica* 51(2): 507-512.

Lillard, Lee et Hong Tan. 1992. "Private Sector Training: Who Gets It and What Are Its Effects? " In Ronald Ehrenberg, ed., *Research in Labor Economics, Volume 13*. Greenwich, CT: J.A.I. Press, p. 1-62.

Michalopoulos, Charles, David Card, Lisa Gennetian, Kristen Harknett et Philip Robins. 2000. *Le Projet d'autosuffisance après trente-six mois : Effets d'un incitatif financier sur l'emploi et le revenu*. Ottawa : Société de recherche sociale appliquée.

Mincer, Jacob. 1974. *Schooling, Experience and Earnings*. New York: Columbia University Press for National Bureau of Economic Research.

Park, Norm, Bob Power, W. Craig Riddell et Ging Wong. 1996. "An Assessment of the Impact of Government-Sponsored Training." *Canadian Journal of Economics*. 29(Special Issue, Part 1). S93-S98.

Puhani, Patrick. 2000. "The Heckman Correction for Sample Selection and Its Critique: A Short Survey." *Journal of Economic Surveys*. 14(1): 53-68.

Riddell, Craig. 1991. « L'évaluation des programmes pour l'emploi et des mesures sociales : Le point sur une question complexe. » Chapitre 2 dans *Évaluation des programmes de main-d'œuvre et de formation : l'expérience de l'Amérique du Nord*. Paris: Organisation de coopération et de développement économiques, p. 49-81.

Rosenbaum, Paul et Donald Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*. 70(1): 41-55.

Smith, Jeffrey. 2000. "A Critical Survey of Empirical Methods for Evaluating Active Labor Market Policies." *Swiss Journal of Economics and Statistics*. 136(3):247-268.

Smith, Jeffrey, et Arthur Sweetman. 2001. « Comment améliorer l'évaluation des programmes d'emploi et de formation au Canada. » Document présenté à la conférence « De la théorie à la pratique » - Conférence internationale sur les méthodes d'évaluation, parrainée par la Direction de l'évaluation et de l'exploitation des données, Développement des ressources humaines Canada, Ottawa, les 16 et 17 novembre 2001. Adresse Internet: http://www11.hrdc-drhc.gc.ca/edd-doc/fttp/smith_sweetman_f.shtml (accédé le 3 juillet, 2003).

Smith, Jeffrey et Petra Todd. 2003. "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics*. Forthcoming.

Smith, Jeffrey et Petra Todd. 2003. "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics*. Forthcoming.

Trott, Charles, John Baj et Kathleen Shankman. 1993. "An Analysis of Repeating in JTPA in Illinois." Report Prepared for the Illinois Department of Commerce and Community Affairs. DeKalb, IL: Center for Governmental Studies, Northern Illinois University.

Waite, Linda et Maggie Gallagher. 2001. The Case for Marriage: *Why Married People are Happier, Healthier, and Better Off Financially*. New York: Broadway Books.

Warburton, William et Warburton, Rebecca. 2002. "Measuring the Performance of Government Training Programs." C.D. Howe Institute Commentary No. 165. Toronto: C.D. Howe Institute.

Notes

1 Ces questions sont documentées, par exemple, dans Barron, Berger et Black (1997).

2 Voir Heckman, LaLonde et Smith (1999) ainsi qu'Angrist et Krueger (1999) pour des résumés récents.

3 Les données de mars 1998 visent uniquement les résidents du Québec, l'enquête de janvier n'ayant pu être réalisée dans la province en raison d'une tempête de verglas.

4 Taux de réponse pour les cinq groupes de renouvellement utilisés dans le cadre de l'EEFA.

5 Voir Heckman, Smith et Clements (1997) de même que Heckman et Smith (1998) pour une discussion détaillée des effets hétérogènes du traitement. Pour une importante analyse antérieure, voir Björklund et Moffitt (1987).

6 Voir Heckman, Ichimura et Todd (1997, 1998) pour une analyse plus poussée de la question.

7 Voir l'analyse de Heckman, Ichimura, Smith et Todd (1998).

8 Les estimations des autres paramètres sont présentées au tableau A.3.

9 Les estimations des autres paramètres sont présentées aux tableaux A.4 à A.6.

Tableaux d'annexe

Tableau A1

Estimations complètes pour le modèle de probabilité linéaire, le modèle probit et les modèles de VI du Tableau 7

Coefficients (x100) et effets marginaux du probit (x100) estimés

	MPL						Probit						VI					
	Hommes			Femmes			Hommes			Femmes			Hommes			Femmes		
Terre-Neuve	-23,52	(2,59)	***	-15,80	(2,52)	***	-26,12	(3,15)	***	-17,45	(2,95)	***	-17,29	(3,28)	***	-1,17	(5,15)	
Île-du-Prince-Édouard	-12,55	(3,06)	***	-5,07	(2,82)	*	-13,61	(3,44)	***	-5,28	(3,20)	*	-7,88	(3,60)	**	5,67	(5,02)	
Nouvelle-Écosse	-7,42	(1,99)	***	-6,05	(2,05)	***	-8,42	(2,35)	***	-6,92	(2,41)	***	-7,52	(2,44)	***	-0,91	(3,70)	
Nouveau-Brunswick	-11,35	(2,11)	***	-10,46	(2,10)	***	-12,46	(2,49)	***	-11,95	(2,46)	***	-6,85	(2,60)	***	-0,71	(4,11)	
Québec	-5,27	(1,50)	***	-2,89	(1,63)	*	-6,01	(1,70)	***	-3,36	(1,87)	*	1,39	(2,32)		17,38	(5,43)	***
Manitoba	7,03	(1,59)	***	2,79	(1,99)		7,19	(1,47)	***	2,95	(2,20)		8,28	(2,04)	***	3,51	(3,59)	
Saskatchewan	4,81	(1,74)	***	4,60	(1,91)	**	4,78	(1,72)	***	5,18	(2,14)	**	5,60	(2,27)	***	4,61	(3,54)	
Alberta	5,33	(1,49)	***	4,54	(1,86)	**	6,06	(1,57)	***	5,11	(2,08)	**	7,36	(2,09)	***	5,17	(3,34)	
Colombie-Britannique	-6,19	(1,77)	***	-4,62	(1,95)	**	-7,01	(2,14)	***	-5,48	(2,31)	**	-5,97	(2,13)	***	-5,88	(3,56)	*
Région métropolitaine de recensement	-1,24	(1,39)		0,91	(1,58)		-1,45	(1,42)		0,94	(1,75)		-2,03	(1,78)		1,22	(2,64)	
Toronto/Montréal/Vancouver	4,01	(1,44)	***	4,71	(1,60)	***	4,35	(1,51)	***	5,51	(1,79)	***	7,18	(1,94)	***	2,85	(2,94)	
Centre urbain	-1,56	(1,98)		1,91	(1,99)		-1,62	(2,05)		2,09	(2,16)		-2,80	(2,31)		-1,22	(3,51)	
Région rurale	-2,67	(1,58)	*	-1,56	(1,70)		-2,94	(1,63)	*	-1,80	(1,88)		-3,03	(1,83)	*	-0,18	(2,83)	
Âge	5,60	(0,46)	***	5,00	(0,44)	***	4,77	(0,46)	***	5,20	(0,51)	***	4,94	(0,66)	***	2,84	(1,08)	***
Âge au carré	-0,07	(0,01)	***	-0,06	(0,01)	***	-0,06	(0,01)	***	-0,07	(0,01)	***	-0,06	(0,01)	***	-0,03	(0,01)	***
Études secondaires partielles	13,08	(2,93)	***	3,75	(2,94)		8,76	(2,66)	***	3,15	(3,04)		14,12	(3,13)	***	6,43	(3,75)	*
11e - 13e année terminée	8,71	(2,16)	***	13,03	(2,09)	***	8,38	(2,32)	***	12,66	(2,19)	***	4,96	(2,61)	*	-0,26	(4,34)	
Études postsecondaires partielles	-1,82	(2,28)		3,44	(2,53)		-1,79	(2,36)		3,15	(2,70)		-7,29	(3,19)	**	-5,74	(4,99)	
Certificat ou diplôme postsecondaire	2,80	(2,07)		5,20	(2,33)	**	3,04	(2,18)		6,17	(2,55)	***	2,27	(2,71)		-4,84	(4,90)	
Baccalauréat/maîtrise/doctorat	6,16	(1,22)	***	3,91	(1,55)	**	7,40	(1,39)	***	5,76	(2,03)	***	-0,85	(2,39)		-15,51	(5,64)	***
Né à l'étranger	-2,22	(1,61)		-7,06	(1,83)	***	-3,25	(1,86)	*	-8,54	(2,12)	***	-0,15	(2,07)		6,31	(4,78)	
Lieu de naissance manquant	-27,03	(10,40)	***	4,70	(5,86)		-33,35	(11,08)	***	7,51	(7,53)		-31,12	(11,65)	***	15,22	(18,12)	
Tous les types de formation	7,82	(1,04)	***	15,91	(1,23)	***	9,15	(1,18)	***	18,60	(1,38)	***	65,56	(12,74)	***	163,06	(34,59)	***
Constante	-40,24	-9,69	***	-45,30	(9,29)	***							-42,23	(12,07)	***	-35,13	(17,80)	**
Observations	10 469			11 938			10 469			11 938			10 469			11 938		
Log-vraisemblance							-4414,4			-6541								
R-carré	0,17			0,17														
Probabilité observée (à la moyenne)							0,80			0,67								
Probabilité prévue (à la moyenne)							0,83			0,69								

* Coefficients statistiquement significatifs à 10 %.
** Coefficients statistiquement significatifs à 5 %.
*** Coefficients statistiquement significatifs à 1 %.

Notes : Les écarts-types estimés (robustes) apparaissent entre parenthèses.

Les coefficients présentés sont pondérés à l'aide des facteurs de pondération fournis par Statistique Canada pour tenir compte de l'échantillon stratifié et de la non-réponse.

Tableau A2

Estimations complètes pour le modèle probit bivarié du Tableau 7

Effets marginaux du probit estimés (x100)

	Probit bivarié							
	Emploi				Incidence de la formation			
	Hommes		Femmes		Hommes		Femmes	
Terre-Neuve	-0,46	(0,09) ***	-0,29	(0,07) ***	-0,37	(0,09) ***	-0,35	(0,08) ***
Île-du-Prince-Édouard	-0,23	(0,10) **	-0,04	(0,08)	-0,25	(0,10) **	-0,23	(0,09) ***
Nouvelle-Écosse	-0,23	(0,07) ***	-0,12	(0,06) **	0,02	(0,07)	-0,10	(0,07)
Nouveau-Brunswick	-0,21	(0,07) ***	-0,20	(0,06) ***	-0,22	(0,07) ***	-0,20	(0,07) ***
Québec	0,01	(0,06)	0,08	(0,06)	-0,38	(0,06) ***	-0,45	(0,06) ***
Manitoba	0,30	(0,07) ***	0,08	(0,06)	-0,06	(0,07)	-0,01	(0,07)
Saskatchewan	0,18	(0,07) **	0,13	(0,06) **	-0,06	(0,07)	0,02	(0,06)
Alberta	0,26	(0,07) ***	0,13	(0,06) **	-0,11	(0,06) *	-0,02	(0,06)
Colombie-Britannique	-0,20	(0,07) ***	-0,14	(0,06) **	0,01	(0,06)	0,01	(0,06)
Région métropolitaine de recensement	-0,07	(0,05)	0,03	(0,05)	0,05	(0,05)	-0,01	(0,05)
Toronto/Montréal/Vancouver	0,23	(0,06) ***	0,12	(0,05) **	-0,15	(0,06) **	0,04	(0,06)
Centre urbain	-0,08	(0,07)	0,03	(0,06)	0,05	(0,07)	0,08	(0,07)
Région rurale	-0,10	(0,05) *	-0,03	(0,05)	0,01	(0,06)	-0,01	(0,06)
Âge	0,14	(0,02) ***	0,11	(0,02) ***	0,00	(0,02)	0,05	(0,02) ***
Âge au carré	0,00	(0,00) ***	0,00	0,00 ***	0,00	(0,00)	0,00	(0,00) ***
Études secondaires partielles	0,31	(0,08) ***	0,12	(0,08)	-0,05	(0,11)	0,12	(0,15)
11e - 13e année terminée	0,13	(0,08) *	0,20	(0,06) ***	0,26	(0,07) ***	0,45	(0,07) ***
Études postsecondaires partielles	-0,21	(0,09) **	0,00	(0,08)	0,28	(0,08) ***	0,17	(0,08) **
Certificat ou diplôme postsecondaire	0,07	(0,08)	0,07	(0,07)	0,03	(0,08)	0,22	(0,07) ***
Baccalauréat/maîtrise/doctorat	0,07	(0,07)	-0,01	(0,06)	0,30	(0,06) ***	0,33	(0,06) ***
Né à l'étranger	-0,04	(0,06)	-0,09	(0,06)	-0,16	(0,06) **	-0,30	(0,06) ***
Lieu de naissance manquant	-0,85	(0,27) ***	0,27	(0,27)	0,21	(0,30)	-0,18	(0,33)
Présence d'un conjoint					0,44	(0,07) ***	0,07	(0,07)
Conjoint – Études secondaires partielles					0,08	(0,16)	0,19	(0,11) *
Conjoint – 11e - 13e année terminée					0,13	(0,08) *	0,02	(0,08)
Conjoint – Études postsecondaires partielles					0,03	(0,08)	-0,01	(0,10)
Conjoint – Certificat ou diplôme postsecondaire					-0,01	(0,08)	0,06	(0,09)
Conjoint – Baccalauréat ou plus					0,15	(0,06) **	0,02	(0,06)
1 enfant					0,13	(0,11)	-0,04	(0,09)
2 enfants et plus					0,20	(0,19)	-0,08	(0,10)
1 enfant et un conjoint					-0,04	(0,13)	-0,03	(0,11)
2 enfants et plus et un conjoint					-0,14	(0,20)	0,05	(0,12)
1 enfant d'âge préscolaire					-0,22	(0,24)	-0,32	(0,11) ***
2 enfants et plus d'âge préscolaire					-0,16	(0,08) *	-0,27	(0,08) ***
Enfants d'âge préscolaire et conjoint					0,09	(0,24)	0,13	(0,12)
Tous les types de formation	1,67	(0,09) ***	1,64	(-0,12) ***				
Constante	-2,53	-0,38 ***	-2,38	-0,32 ***	-1,29	-0,42 ***	-2,20	(0,39) ***
Observations	11 469		11 938					
Log-vraisemblance	-7207403		-8007787,8					
Rho	-0,85		-0,71					
Pr(Rho=0)	0,00		0,00					

* Coefficients statistiquement significatifs à 10 %.
** Coefficients statistiquement significatifs à 5 %.
*** Coefficients statistiquement significatifs à 1 %.

Notes : Les écarts-types estimés (robustes) apparaissent entre parenthèses.

Les coefficients présentés sont pondérés à l'aide des facteurs de pondération fournis par Statistique Canada pour tenir compte de l'échantillon stratifié et de la non-réponse.

Tableau A3

Estimations complètes pour le modèle de probabilité linéaire, le modèle probit et les modèles de VI des Tableaux 8A et 8B

Coefficients (x100) et effets marginaux du probit (x100) estimés

	MPL						Probit						VI					
	Hommes			Femmes			Hommes			Femmes			Hommes			Femmes		
Terre-Neuve	-23,00	(2,58)	***	-15,25	(2,53)	***	-25,12	(3,17)	***	-16,66	(3,02)	***	-18,71	(3,82)	***	2,02	(6,44)	
Île-du-Prince-Édouard	-11,75	(3,05)	***	-4,99	(2,81)	*	-11,98	(3,35)	***	-5,14	(3,23)		-5,33	(4,37)		6,03	(5,53)	
Nouvelle-Écosse	-7,67	(1,99)	***	-5,83	(2,05)	***	-8,85	(2,41)	***	-6,77	(2,46)	***	-6,75	(3,56)	*	2,58	(4,84)	
Nouveau-Brunswick	-10,89	(2,10)	***	-10,08	(2,07)	***	-11,41	(2,45)	***	-11,44	(2,50)	***	-7,30	(3,25)	**	2,61	(4,75)	
Québec	-5,14	(1,48)	***	-2,47	(1,62)		-5,68	(1,64)	***	-2,94	(1,85)		-0,42	(3,54)		21,23	(5,92)	***
Manitoba	6,95	(1,59)	***	3,20	(1,98)		6,85	(1,40)	***	3,64	(2,14)	*	9,92	(3,51)	***	5,18	(4,65)	
Saskatchewan	4,47	(1,73)	***	4,70	(1,88)	**	4,39	(1,68)	***	5,54	(2,07)	***	2,81	(3,12)		7,04	(4,26)	*
Alberta	5,20	(1,49)	***	4,59	(1,84)	**	5,68	(1,52)	***	5,09	(2,05)	**	6,48	(2,72)	**	6,73	(3,74)	*
Colombie-Britannique	-5,96	(1,75)	***	-4,14	(1,93)	**	-6,56	(2,07)	***	-4,90	(2,29)	**	-1,45	(3,08)		-3,47	(4,02)	
Région métropolitaine de recensement	-1,05	(1,38)		0,92	(1,57)		-1,10	(1,38)		1,03	(1,76)		-0,10	(2,18)		1,98	(3,12)	
Toronto/Montréal/Vancouver	4,20	(1,42)	***	4,59	(1,59)	***	4,50	(1,42)	***	5,56	(1,77)	***	6,14	(2,41)	**	2,46	(3,30)	
Centre urbain	-1,33	(1,97)		2,41	(2,00)		-1,42	(2,00)		2,62	(2,15)		0,89	(2,95)		2,25	(4,31)	
Région rurale	-2,57	(1,58)		-1,63	(1,69)		-2,80	(1,60)	*	-1,90	(1,90)		-2,57	(2,28)		0,73	(3,61)	
Âge	5,35	(0,45)	***	4,69	(0,44)	***	4,28	(0,43)	***	4,71	(0,49)	***	2,25	(1,06)	**	1,66	(1,85)	
Âge au carré	-0,07	(0,01)	***	-0,06	(0,01)	***	-0,06	(0,00)	***	-0,06	(0,01)	***	-0,04	(0,01)	***	-0,02	(0,02)	
Études secondaires partielles	13,19	(2,92)	***	4,16	(2,92)		8,53	(2,60)	***	3,49	(2,97)		13,17	(3,70)	***	8,62	(3,88)	**
11e - 13e année terminée	8,11	(2,11)	***	12,52	(2,06)	***	7,34	(2,16)	***	11,54	(2,15)	***	2,87	(2,98)		-2,70	(4,80)	
Études postsecondaires partielles	-1,75	(2,24)		3,71	(2,44)		-1,69	(2,26)		3,61	(2,62)		-0,57	(3,74)		-4,26	(5,08)	
Certificat ou diplôme postsecondaire	2,72	(2,03)		4,97	(2,25)	**	2,94	(2,09)		5,75	(2,48)	**	0,33	(3,08)		-8,10	(5,77)	
Baccalauréat/maîtrise/doctorat	6,03	(1,21)	***	2,98	(1,55)	*	7,20	(1,34)	***	4,57	(2,08)	**	7,85	(3,90)	**	-22,07	(6,86)	***
Né à l'étranger	-2,09	(1,59)		-5,90	(1,79)	***	-3,09	(1,78)	*	-7,00	(2,06)	***	1,37	(2,67)		12,05	(6,29)	*
Lieu de naissance manquant	-25,29	(10,66)	**	4,12	(6,08)		-31,70	(12,11)	***	6,85	(7,58)		-32,17	(14,33)	**	12,08	(21,08)	
Formation financée par l'employeur	12,49	(0,81)	***	25,24	(0,98)	***	15,31	(0,89)	***	31,28	(1,07)	***	70,61	(16,65)	***	188,60	(64,36)	***
Formation financée par l'État	-16,14	(5,12)	***	-17,91	(4,33)	***	-17,43	(5,75)	***	-19,30	(4,60)	***	-40,98	(141,07)		-72,77	(110,33)	
Formation autofinancée	-0,35	(2,48)		4,09	(1,96)	**	-1,44	(2,71)		4,30	(2,24)	*	-132,70	(54,02)	**	156,43	(106,65)	
Constante	-34,74	-9,49	***	-38,30	(9,13)	***							29,19	-21,13		-8,40	(34,75)	
Observations	10 469			11 938			10 469			11 938			10 469			11 938		
Log-vraisemblance							-4317,1506			-6314,386								
R-carré	0,18			0,19														
Probabilité observée							0,80			0,67								
Probabilité prévue (à la moyenne)							0,84			0,71								

* Coefficients statistiquement significatifs à 10 %.
** Coefficients statistiquement significatifs à 5 %.
*** Coefficients statistiquement significatifs à 1 %.

Notes : Les écarts-types estimés (robustes) apparaissent entre parenthèses.

Les coefficients présentés sont pondérés à l'aide des facteurs de pondération fournis par Statistique Canada pour tenir compte de l'échantillon stratifié et de la non-réponse.

N° 81-595-MIF2003008 au catalogue

Tableau A4

Estimations complètes pour le modèle linéaire et le modèle des VI du Tableau 9

Coefficients estimés (x100)

	Modèle linéaire						Variables instrumentales					
	Hommes			Femmes			Hommes			Femmes		
Terre-Neuve	-194,2	(22,9)	***	-108,6	(12,3)	***	-95,6	(36,2)	***	-33,9	(27,8)	
Île-du-Prince-Édouard	-185,1	(22,7)	***	-39,9	(15,8)	**	-111,2	(38,0)	***	14,9	(27,2)	
Nouvelle-Écosse	-125,4	(19,5)	***	-70,5	(12,9)	***	-127,1	(29,8)	***	-44,3	(20,5)	**
Nouveau-Brunswick	-90,2	(19,7)	***	-53,4	(13,1)	***	-19,0	(30,2)		-3,7	(23,4)	
Québec	-65,8	(18,2)	***	-52,8	(12,1)	***	39,5	(30,8)		50,6	(32,0)	
Manitoba	-45,9	(19,3)	**	-58,5	(12,2)	***	-26,2	(28,4)		-54,8	(19,6)	***
Saskatchewan	-97,1	(19,7)	***	-51,6	(12,3)	***	-84,5	(28,2)	***	-51,6	(20,0)	***
Alberta	-31,6	(20,7)		-35,7	(12,9)	***	0,6	(29,0)		-32,5	(19,4)	*
Colombie-Britannique	-55,9	(21,2)	***	-40,5	(15,2)	***	-52,3	(28,0)	*	-47,0	(21,0)	**
Région métropolitaine de recensement	46,8	(16,1)	***	28,4	(10,6)	***	34,4	(22,1)		30,0	(15,4)	*
Toronto/Montréal/Vancouver	-1,4	(19,1)		60,8	(12,8)	***	48,8	(27,3)	*	51,3	(17,6)	***
Centre urbain	-4,8	(21,1)		12,0	(13,5)		-24,4	(29,1)		-4,0	(20,6)	
Région rurale	-91,9	(16,3)	***	-48,4	(10,5)	***	-97,5	(22,1)	***	-41,4	(16,0)	***
Âge	63,0	(4,5)	***	34,2	(3,2)	***	52,5	(6,6)	***	23,2	(5,8)	***
Âge au carré	-0,7	(0,1)	***	-0,4	(0,0)	***	-0,6	(0,1)	***	-0,3	(0,1)	***
Études secondaires partielles	81,2	(20,2)	***	-0,4	(12,9)		97,7	(27,5)	***	13,3	(17,6)	
11e - 13e année terminée	72,4	(19,2)	***	70,4	(10,9)	***	13,1	(28,3)		2,6	(24,4)	
Études postsecondaires partielles	-14,8	(26,6)		14,9	(15,4)		-101,3	(39,4)	**	-31,9	(27,4)	
Certificat ou diplôme postsecondaire	64,7	(25,0)	***	49,4	(15,6)	***	56,3	(33,6)	*	-1,8	(27,8)	
Baccalauréat/maîtrise/doctorat	118,2	(21,4)	***	154,6	(16,8)	***	7,3	(33,3)		55,5	(35,0)	
Né à l'étranger	-60,8	(19,5)	***	-71,4	(14,0)	***	-28,1	(26,0)		-3,2	(28,3)	
Lieu de naissance manquant	-311,4	(57,2)	***	192,4	(83,8)	**	-376,1	(112,8)	***	246,0	(130,2)	*
Tous les types de formation	172,8	(15,8)	***	158,9	(11,4)	***	1 086,5	(188,4)	***	909,5	(205,3)	***
Observations	10 469			11 938			10 469			11 938		
R-carré	0,16			0,23								

* Coefficients statistiquement significatifs à 10 %.

** Coefficients statistiquement significatifs à 5 %.

*** Coefficients statistiquement significatifs à 1 %.

Notes : Les écarts-types estimés (robustes) apparaissent entre parenthèses. Le terme constant estimé est omis.

Les coefficients présentés sont pondérés à l'aide des facteurs de pondération fournis par Statistique Canada pour tenir compte de l'échantillon stratifié et de la non-réponse.

Tableau A5

Estimations complètes pour le modèle normal bivarié du Tableau 9

Coefficients (x100) et effets marginaux probit (x100) estimés

	Gains hebdomadaires ($)						Incidence de la formation					
	Hommes			Femmes			Hommes			Femmes		
Terre-Neuve	-168,9	(25,20)	***	-104,1	(12,30)	***	-0,40	(0,09)	***	-0,37	(0,09)	***
Île-du-Prince-Édouard	-166,1	(25,00)	***	-36,6	(15,70)	**	-0,25	(0,10)	**	-0,24	(0,09)	***
Nouvelle-Écosse	-125,8	(20,50)	***	-68,9	(12,90)	***	0,01	(0,07)		-0,10	(0,07)	***
Nouveau-Brunswick	-71,9	(21,30)	***	-50,4	(13,10)	***	-0,25	(0,07)	***	-0,21	(0,07)	***
Québec	-38,8	(21,10)	*	-46,6	(12,10)	***	-0,40	(0,06)	***	-0,50	(0,06)	***
Manitoba	-40,8	(20,30)	**	-58,3	(12,20)	***	-0,07	(0,07)		0,00	(0,07)	
Saskatchewan	-93,9	(20,30)	***	-51,6	(12,40)	***	-0,05	(0,07)		0,02	(0,06)	
Alberta	-23,3	(21,7)		-35,5	(12,90)	***	-0,13	(0,06)	**	-0,01	(0,06)	
Colombie-Britannique	-54,9	(21,60)	**	-40,9	(15,20)	***	-0,01	(0,07)		0,01	(0,06)	
Région métropolitaine de recensement	43,6	(16,70)	***	28,5	(10,60)	***	0,04	(0,05)		-0,02	(0,05)	
Toronto/Montréal/Vancouver	11,5	(20,4)		60,2	(12,80)	***	-0,16	(0,06)	***	0,05	(0,06)	
Centre urbain	-9,8	(21,9)		11,0	(13,5)		0,06	(0,07)		0,07	(0,07)	
Région rurale	-93,3	(16,70)	***	-48,0	(10,50)	***	0,03	(0,06)		-0,02	(0,06)	
Âge	60,3	(4,80)	***	33,6	(3,20)	***	0,01	(0,02)		0,07	(0,02)	***
Âge au carré	-0,7	(0,10)	***	-0,4	(0,00)	***	0,00	(0,00)		0,00	(0,00)	***
Études secondaires partielles	85,5	(20,90)	***	0,5	(12,9)		-0,05	(0,12)		0,14	(0,14)	
11e - 13e année terminée	57,2	(20,70)	***	66,3	(11,00)	***	0,25	(0,08)	***	0,46	(0,07)	***
Études postsecondaires partielles	-37,0	(28,2)		12,0	(15,5)		0,29	(0,08)	***	0,20	(0,08)	**
Certificat ou diplôme postsecondaire	62,5	(25,50)	**	46,3	(15,70)	***	0,04	(0,08)		0,21	(0,07)	***
Baccalauréat/maîtrise/doctorat	89,8	(23,80)	***	148,6	(17,10)	***	0,27	(0,06)	***	0,32	(0,06)	***
Né à l'étranger	-52,4	(20,10)	***	-67,3	(14,10)	***	-0,12	(0,06)	**	-0,30	(0,06)	***
Lieu de naissance manquant	-328,0	(63,20)	***	195,6	(85,10)	**	0,22	(0,29)		-0,23	(0,35)	
Présence d'un conjoint							0,43	(0,08)	***	0,08	(0,08)	
Conjoint – Études secondaires partielles							0,13	(0,17)		0,20	(0,12)	*
Conjoint – 11e - 13e année terminée							0,17	(0,08)	**	-0,01	(0,08)	
Conjoint – Études postsecondaires partielles							0,11	(0,09)		0,03	(0,10)	
Conjoint – Certificat ou diplôme postsecondaire							-0,09	(0,09)		0,05	(0,10)	
Conjoint – Baccalauréat ou plus							0,20	(0,07)	***	0,06	(0,07)	
1 enfant							0,04	(0,15)		0,00	(0,10)	
2 enfants et plus							0,33	(0,22)		-0,04	(0,11)	
1 enfant et un conjoint							0,09	(0,16)		-0,08	(0,12)	
2 enfants et plus et un conjoint							-0,27	(0,23)		0,03	(0,13)	
1 enfant d'âge préscolaire							-0,33	(0,29)		-0,18	(0,12)	
2 enfants et plus d'âge préscolaire							-0,17	(0,10)	*	-0,25	(0,09)	***
Enfants d'âge préscolaire et un conjoint							0,16	(0,30)		0,10	(0,13)	
Tous les types de formation	407,2	(90,4)	***	204,4	(18,5)	***						
Constant	-994,1	(-102,2)	***	-509,3	(-68,7)	***	-1,47	(-0,45)	***	-2,59	(-0,39)	***
Observations	10 469			11 938								
Log-vraisemblance	-61962067			-58169147								
Rho	-0,32			-0,09								
Pr(Rho=0)	0,0088			0,0009								

* Coefficients statistiquement significatifs à 10 %.
** Coefficients statistiquement significatifs à 5 %.
*** Coefficients statistiquement significatifs à 1 %.

Notes : Les écarts-types estimés (robustes) apparaissent entre parenthèses.

Les coefficients présentés sont pondérés à l'aide des facteurs de pondération fournis par Statistique Canada pour tenir compte de l'échantillon stratifié et de la non-réponse.

N° 81-595-MIF2001008 au catalogue

Tableau A6

Estimations complètes du modèle de régression linéaire et du modèle des VI des Tableaux 10A et 10B

Coefficients estimés (x100)

	Modèle linéaire						Variables instrumentales					
	Hommes			Femmes			Hommes			Femmes		
Terre-Neuve	-186,8	(22,2)	***	-102,3	(12,4)	***	-112,0	(38,0)	***	-10,1	(38,7)	
Île-du-Prince-Édouard	-171,6	(22,5)	***	-38,8	(15,6)	**	-86,0	(41,6)	**	21,3	(32,4)	
Nouvelle-Écosse	-127,6	(19,5)	***	-69,7	(12,6)	***	-114,2	(39,0)	***	-16,5	(29,3)	
Nouveau-Brunswick	-84,1	(19,5)	***	-50,4	(12,8)	***	-33,3	(34,0)		23,4	(28,3)	
Québec	-62,0	(18,0)	***	-49,5	(11,9)	***	35,0	(42,3)		86,3	(37,0)	**
Manitoba	-44,3	(19,2)	**	-53,9	(11,8)	***	1,4	(41,5)		-46,3	(28,1)	*
Saskatchewan	-102,0	(19,5)	***	-51,3	(12,1)	***	-100,4	(37,1)	***	-34,8	(25,1)	
Alberta	-32,8	(20,4)		-35,5	(12,7)	***	-4,6	(33,9)		-21,8	(22,9)	
Colombie-Britannique	-49,6	(20,9)	**	-36,3	(14,8)	**	-6,0	(34,2)		-32,1	(26,1)	
Région métropolitaine de recensement	50,2	(15,9)	***	28,0	(10,7)	***	51,8	(24,2)	**	36,4	(19,3)	*
Toronto/Montréal/Vancouver	1,9	(18,8)		59,7	(12,6)	***	43,6	(30,0)		47,7	(20,9)	**
Centre urbain	0,5	(21,0)		15,6	(13,4)		13,8	(32,2)		19,0	(26,1)	
Région rurale	-30,6	(16,3)	***	-49,8	(10,5)	***	-95,6	(24,3)	***	-32,9	(22,2)	
Âge	57,6	(4,4)	***	31,2	(3,1)	***	24,0	(10,3)	**	15,2	(10,9)	
Âge au carré	-0,7	(0,1)	***	-0,4	(0,0)	***	-0,3	(0,1)	**	-0,2	(0,1)	
Études secondaires partielles	82,1	(20,4)	***	3,4	(12,8)		85,3	(29,4)	***	28,5	(19,7)	
11e - 13e année terminée	62,1	(19,0)	***	65,7	(10,6)	***	-7,7	(30,7)		-19,4	(28,4)	
Études postsecondaires partielles	-11,6	(26,1)		16,8	(14,7)		-41,6	(46,4)		-25,4	(30,6)	
Certificat ou diplôme postsecondaire	62,3	(24,6)	**	48,7	(14,9)	***	38,8	(35,5)		-31,2	(35,0)	
Baccalauréat/maîtrise/doctorat	120,8	(21,2)	***	146,4	(16,6)	***	97,5	(46,3)	**	3,7	(43,0)	
Né à l'étranger	-57,1	(19,2)	***	-60,5	(13,6)	***	-8,9	(30,2)		38,8	(37,5)	
Lieu de naissance manquant	-288,9	(56,1)	***	186,4	(83,2)	**	-367,6	(135,0)	***	233,6	(154,2)	
Formation financée par l'employeur	256,6	(17,1)	***	256,2	(11,9)	***	1 162,5	(203,6)	***	1 112,1	(391,0)	***
Formation financée par l'État	-122,4	(50,5)	**	-93,0	(23,3)	***	720,1	(1 736,9)		-666,9	(681,5)	
Formation autofinancée	-60,5	(28,4)	**	15,1	(19,2)		-969,6	(646,0)		1 035,2	(679,1)	
Observations	10 469			11 938			10 469			11 938		
R-carré	0,18			0,27								

* Coefficients statistiquement significatifs à 10 %.

** Coefficients statistiquement significatifs à 5 %.

*** Coefficients statistiquement significatifs à 1 %.

Notes : Les écarts-types estimés (robustes) apparaissent entre parenthèses.

Les coefficients présentés sont pondérés à l'aide des facteurs de pondération fournis par Statistique Canada pour tenir compte de l'échantillon stratifié et de la non-réponse.

Culture, tourisme et Centre de la statistique de l'éducation

Documents de recherche

Index cumulatif

La **Division de la Culture, du tourisme et Centre de la statistique de l'éducation** de Statistique Canada élabore des enquêtes, fournit des statistiques et effectue des recherches et des analyses sur les questions d'actualité dans ses trois domaines de responsabilité.

Le **Programme de la statistique culturelle** élabore et diffuse des données actuelles et détaillées sur le secteur culturel au Canada. Ce programme gère une douzaine d'enquêtes/recensements périodiques et de banques de données afin de produire des données qui appuient la prise de décisions stratégiques et la gestion des programmes. Les questions d'actualité incluent les incidences économiques de la culture, la consommation de biens et de services culturels, les dépenses culturelles de l'État, des particuliers et des entreprises, le marché du travail du secteur de la culture, ainsi que le commerce international des biens et des services culturels. Des articles analytiques sont diffusés dans la publication de prestige *La culture en perspective* (www.statcan.ca/francais/IPS/Data/87-004-XIF.htm) et dans *Arts, culture et loisirs – Documents de recherche*.

Le **Programme de la statistique du tourisme** fournit des renseignements sur la demande portant sur le tourisme intérieur et international. Le programme couvre l'Enquête sur les voyages des Canadiens (EVC) et l'Enquête sur les voyages internationaux (EVI). Ensemble, ces deux enquêtes donnent des renseignements sur le nombre et les caractéristiques des voyages et des voyageurs en provenance et à destination du Canada et à l'intérieur du pays. Des articles analytiques sont diffusés dans la publication de prestige *Info-voyages* (www.statcan.ca/francais/IPS/Data/87-003-XIF.htm) et dans *Voyages et tourisme – Documents de recherche*.

Le **Centre de la statistique de l'éducation** vise à concevoir et à réaliser un programme complet de collecte et d'analyse de données statistiques pancanadiennes sur l'éducation comme aide aux décisions de politiques et à la gestion des programmes et aussi comme moyen de garantir qu'une information précise et utile sera mise à la disposition du public et des autres intervenants en éducation au Canada. Le Centre mène 15 enquêtes auprès des établissements d'enseignement et plus de 10 enquêtes-ménages sur l'éducation. Des articles analytiques sont diffusés dans la publication de prestige *Revue trimestrielle de l'éducation*(www.statcan.ca/francais/IPS/Data/81-003-XIF.htm, dans divers monographies et dans *Éducation, compétences et apprentissage – Documents de recherche* (www.statcan.ca/francais/IPS/Data/81-595-MIF.htm).

N° 81-595-MIF2003008 au catalogue

Ci-dessous figure un index cumulatif des documents de recherche sur la culture, le tourisme et l'éducation publiés à date.

Arts, culture et loisirs – Documents de recherche

À venir

Voyages et tourisme – Documents de recherche

À venir

Éducation, compétences et apprentissage – Documents de recherche

81-595-MIF2002001	Comprendre l'écart rural-urbain dans le rendement en lecture
81-595-MIF2003002	Services canadiens d'éducation et de formation à l'étranger: le rôle des contrats financés par les institutions financières internationales
81-595-MIF2003003	Trouver sa voie : profil des jeunes diplômés canadiens
81-595-MIF2003004	Étudier, travailler et décrocher : Relation entre le travail pendant les études secondaires et le décrochage scolaire
81-595-MIF2003005	Établir le lien entre les évaluations provinciales des élèves et les évaluations nationales et internationales
81-595-MIF2003006	Qui poursuit des études postsecondaires et à quel moment : Parcours choisis par les jeunes de 20 ans
81-595-MIF2003007	Accès, persévérance et financement : Premiers résultats de l'Enquête sur la participation aux études postsecondaires (EPÉP)
81-595-MIF2003008	L'incidence de l'éducation et de la formation des adultes sur la situation sur le marché du travail au Canada